

# Impacts of Labor Market Programs

## *Comparison of Experimental and Non-experimental Methods*

Ana Teresa Muñoz Delgado



Master Thesis at the Department of Economics  
Faculty of Social Sciences  
Master of Philosophy in Economics

UNIVERSITETET I OSLO

May 14, 2012





Impacts of Labor Market Programs:  
*Comparison of Experimental and Non-  
experimental Methods*

© Ana Teresa Muñoz Delgado

2012

Impacts of Labor Market Programs: Comparison of Experimental and Non-experimental  
Methods

Ana Teresa Muñoz Delgado

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

# Abstract

This study compares experimental and non-experimental estimators using randomized data from the National Evaluation of Welfare-to-Work Strategies. Our main question is: can non-experimental methods match results obtained using random assignment? There are three key empirical conclusions from this study. First, results obtained using non-experimental data can lead to wrong conclusions about the causal effects of a training program. Second, biases obtained from non-experimental data depend not only on the econometric procedure used but also on the chosen comparison group. Third, comparisons longer ahead in time are more susceptible to selection bias problems. In other words, medium-run bias was larger than short-run bias.



# Preface

This Master Thesis is submitted for the degree of Master in Philosophy in Economics at the University of Oslo.

I would like to express my gratitude to the Microeconomics Research Department at Statistics Norway, where this thesis was written as part of a paid engagement. In particular, I want to thank my supervisor Tarjei Havnes, for his guidance, insightful comments and patience. I also want to thank Magne Mogstad for his helpful insights and ideas.

Lastly, I would like to show my gratitude to all of those who supported me in any respect; my partner, family, friends and colleagues.



# Table of Contents

1	Introduction .....	1
2	Context .....	4
2.1	Context.....	4
2.2	National Evaluation of Welfare-to-Work Strategies (NEWWS) .....	7
2.2.1	Labor Force Attachment Approach (LFA).....	8
2.2.2	Human Capital Development Approach (HCD) .....	9
2.2.3	Portland’s employment-focused program .....	11
2.2.4	Participation patterns for treatment group members .....	12
2.2.5	Participation patterns for control group members .....	15
2.3	Literature Review .....	15
3	Empirical Strategy.....	17
3.1	Random Assignment.....	19
3.2	Selection on observables .....	22
3.2.1	Semi-parametric and non-parametric estimation .....	25
3.3	Selection on unobservables .....	29
3.3.1	Differences-in-Differences (DD) .....	29
4	Data and Descriptive Statistics.....	32
5	Implementation.....	37
5.1	Random assignment.....	37
5.2	Ordinary Least Squares (OLS) .....	39
5.3	Propensity score matching (PSM) .....	40
5.4	Differences-in-Differences (DD).....	43
6	Results .....	44
6.1	Experimental results .....	44
6.2	Non-experimental results.....	47
6.3	Estimated bias arising from non-experimental data .....	50
7	Concluding Remarks .....	53
	References .....	54
	Acronyms and Abbreviations.....	57

Appendix A .....	59
Appendix B .....	69
Appendix C .....	73

## List of Tables and Figures

Table 1.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Portland. ....	35
Table 2.- Selected characteristics of female control group members with two years of earnings data prior to random assignment by site.....	36
Table 3.- Results obtained using experimental data. Impacts are measured through the difference of means between the randomized treatment group and the randomized control group.....	45
Figure 1.- Mean quarterly earnings: Portland .....	46
Figure 2.- Mean quarterly earnings: Grand Rapids.....	46
Figure 3.- Mean quarterly earnings: Riverside.....	47
Table 4.- Comparison between short-run and medium-run impacts measured using experimental and non-experimental data: Portland.....	49
Figure 4.- Mean quarterly earnings of control group members in Grand Rapids, Portland and Riverside.....	51
Table 5.- Estimated short-run and medium-run bias for comparisons between control group members in Grand Rapids, Portland and Riverside .....	52
Appendix Figure A.1.- Labor Force Attachment Activities Sequence .....	60
Appendix Figure A.2.- Human Capital Development Activities Sequence.....	60
Appendix Table A.1.- Main characteristics of the training programs by Site .....	61
Appendix Table A.2.- Main characteristics of the training programs.....	63
Appendix Figure A.3.- Labor Force Attachment and Human Capital Development Participation Patterns.....	66
Appendix Table A.3.- Main characteristics by program, treatment group versus control group .....	67
Appendix Figure A.4.- Control Group Participation Patterns.....	68
Appendix Table B.1.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Grand Rapids. ....	70
Appendix Table B.2.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Riverside in need of education.....	71
Appendix Table B.3.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Riverside not in need of education.....	72
Appendix Table C.1.- Comparison between short-run and medium-run impacts measured using experimental and non-experimental data: Grand Rapids .....	74
Appendix Table C.2.- Comparison between short-run and medium-run impacts measured using experimental and non-experimental data: Riverside .....	75





# 1 Introduction

Unobserved heterogeneity and endogeneity permeates the questions that we pose as economists and social scientists. In a seminal paper, LaLonde (1986) used a randomized experiment to evaluate the empirical models used to estimate the impact of labor market programs. His results revealed that many of the econometric procedures used at that time to evaluate employment and training programs would not have yielded accurate or precise estimates of the impact as observed in randomized trials. In particular, even when the econometric estimates pass conventional specification tests, they still fail to replicate experimental results. LaLonde suggests that policymakers should be aware that non-experimental evaluations may contain large and unknown biases resulting from specification errors. The study had a profound impact on empirical economics, and since it was published a large body of theoretical and empirical work has developed methods that aim to more effectively eliminate or reduce the biases created by specification errors. Studies similar to LaLonde (1986) have since been used to evaluate the biases inherent in these methods, finding that non-experimental estimates are often sensitive both to the particular analysis sample and to the econometric procedure.

In the last decades, different econometric methods have been developed in order to meet the requirements of program evaluation. While randomized trials keep being the first best solution when calculating impacts of training programs, researchers have also developed econometric procedures that control for observable and unobservable characteristics. These methods have been particularly useful since randomized data are not easy available. However, it is also important to evaluate and compare the performance of these different procedures in order to identify their strengths and weaknesses. Along these lines, our goal is to reevaluate LaLonde's results (1986) using data from the National Evaluation of Welfare-to-Work Strategies.

In this study, we start by reviewing the context where welfare-to-work reforms arose in the United States. Under the Aid to Families with Dependent Children<sup>1</sup>, a federally funded cash welfare program was created in order to protect children who were deprived of parental support. After the Family Support Act was passed in 1988, permanent welfare payments were

---

<sup>1</sup> In 1935, the Social Security act established a state-level grant called Aid to Families with Dependent Children. This program was in effect until 1996.

replaced by temporary assistance, where parents were expected to be the main supporters of their children. In addition, a mutual obligation between states and welfare-recipients was created. States were responsible for providing support to help welfare recipients to find employment, while recipients were responsible for taking jobs and participating in the Job Opportunities and Basic Skills Training program.

The National Evaluation of Welfare-to-Work Strategies was a study undertaken in seven locales, in order to measure the impacts of these mandatory welfare-to-work programs created following the guidelines imposed by JOBS. At each site, individuals were randomized into a control or treatment group, and their outcomes were followed for a period of five years after random assignment. Control group members were not subject to sanctions and could participate in any activities or services that were available before the implementation of the program. Treatment group members were expected to participate in the activities assigned by the program, at the risk of losing welfare payments if they did not participate.

Our empirical analysis uses NEWWS data covering female sample members with two years of earnings data prior to random assignment, in three sites: Grand Rapids, Portland and Riverside. In order to measure the impacts of welfare-to-work programs on earnings, we use experimental and non-experimental methods. Specifically, we start by calculating the difference of means by treatment status using randomized data. In this approach, average earnings for control group members are compared to average earnings for treatment group members, within the same site. Following LaLonde (1986), we then use alternative econometric models to estimate the impact of labor market programs, where we use non-experimental features of the data. By comparing control group members from one site (control group) to control group members from another site (comparison group), we may evaluate the bias that would arise from non-randomized comparisons. First, we consider selection on observable characteristics, using regression methods and propensity-score matching. Second, we consider selection on unobservables by using a differences-in-differences estimator.

Selection on observables is obtained in both methods by including covariates reflecting earnings prior to randomization, employment prior to randomization and background characteristics. In propensity-score matching, we follow the subclassification method employed by Michalopoulos et al. (2004). Individuals are divided in subgroups with similar characteristics based on their propensity scores and then, a bias is calculated from each

comparison between the control group and the comparison group. The total bias arising from the comparison between the two sites is therefore a weighted average of the biases from each subgroup.

Differences-in-differences estimators are used in order to control for selection on unobservables. The main goal is to estimate the causal effect of program participation on earnings without considering possible fixed effects that are not related to the treatment, also known as common trends.

We might summarize the insights from our research with three main conclusions. First, results obtained using non-experimental data can lead to wrong conclusions about the causal effects of a training program. Second, biases obtained from non-experimental data depend not only on the econometric procedure used but also on the chosen comparison group. Third, comparisons longer ahead in time are more susceptible to selection bias problems. In other words, medium-run bias was larger than short-run bias.

As a concluding remark and a recommendation for future research, our study suggests the potential of the evaluation and comparison of distributional estimators.

The outline of the thesis is as follows. Section 2 reviews the context where all these reforms arose, explains the characteristics of the different training programs and reviews previous literature. Section 3 outlines the empirical strategies. Section 4 describes the data and sample used. Section 5 explains the implementation of the econometric procedures. Section 6 presents the results obtained and section 7 considers the concluding remarks.

## 2 Context

The first part of this section reviews the political context where welfare reform programs arose in the United States. Starting in 1935, the main lines of Aid to Families with Dependent Children (AFDC) program will be explained. Then, we will continue by explaining the main reforms imposed to AFDC by the Family Support Act of 1988. To finish this subsection, we will talk about the main characteristics of the Jobs Opportunities and Basic Skills Training (JOBS) program and its link to the National Evaluation of Welfare-to-Work Strategies (NEWWS) and our research.

The second part explains in more detail the National Evaluation of Welfare-to-Work Strategies and the two main approaches used: Labor Force Attachment and Human Capital Development. In addition, participation patterns for control and treatment group members will be explained.

The third part of this section reviews the previous literature on the topic.

### 2.1 Context<sup>2</sup>

In 1935, the Social Security act established a state-level grant program called Aid to Families with Dependent Children (AFDC). This program aimed to protect children who were deprived of parental support, through the provision of cash welfare payments. Absence of one of the parents, incapacitation, death or unemployment, were some of the reasons leading to welfare support. The AFDC became the major federally funded cash welfare program in the United States and it was in effect until 1996, under the administration of the Department of Health and Human Services.

Later on, welfare reform legislation was passed by the US Congress in 1988. The Family Support Act (FSA) revised the AFDC program to emphasize work, child support and family benefits.

The basic entitlement nature of the AFDC program was not modified by the FSA; however the main focus was on shifting balance from permanent to temporary income maintenance.

---

<sup>2</sup> The discussion in this section is based largely on Gueron (1990), Gueron (1991) and Office of Human Services Policy (1998).



According to the FSA, parents should be the main supporters of their children and government assistance should encourage grant recipients to move off welfare. A mutual obligation between recipients and government was established.

On the one hand, states were responsible for providing incentives and support services to help welfare recipients to find employment. On the other hand, recipients were responsible for taking jobs and participating in the Job Opportunities and Basic Skills Training (JOBS) program created under the FSA.

States were expected to guarantee support services as child care, if those were necessary for an individual's employment or education. Furthermore, the participation in governmental programs could not be mandatory if these services were not provided.

States were also required to provide transitional benefits such as Transitional Child Care (TCC) and Transitional Medicaid Assistance (TMA), in order to help former welfare recipients in their transition to self-support. The TCC and TMA programs provided 12 months of supportive services to former recipients complying with two requirements. Firstly, individuals were only eligible for transitional services if they lost benefits because eligibility limits were exceeded. Some of the reasons leading the income to exceed the limits were: an increase in salary, an increase in the number of hours worked or the expiration of an income disregard. For example, earnings of AFDC recipients were subject to earned income disregards for a maximum of one year, when calculating the welfare grant. Secondly, individuals were required to have received assistance for at least three out of the previous six months prior to losing benefits.

Some other related AFDC amendments stated by the FSA were: increased disregards for earned income and increased disregards for child care. Earned income disregards were defined as the amount of monthly earned income an AFDC recipient may keep before the size of the AFDC benefit is reduced. Disregards for child care were considered as the amount of earnings that could be set aside for child care before the size of the grant is reduced.

As mentioned, recipients were responsible for participating in Job Opportunities and Basic Skills Training Program (JOBS), created under the Family Support Act of 1988. More specifically, welfare recipients were provided with education and job search activities, and required to participate as a condition of receiving AFDC grants.

JOBS was a training program designed to increase welfare recipients' job skills and opportunities. There were two main variations of JOBS: Labor Force Attachment which emphasized the rapid acquisition of employment and Human Capital Development which promoted longer-term education and job training.

The Job Opportunities and Basic Skills Training Program (JOBS) required most single parents with children ages one to five, to enroll in welfare-to-work programs. Moreover, the enrollees were required to participate in these activities for as long as they received welfare grants and were eligible for services.

States were allowed to design their own programs and JOBS would provide federal matching funds for welfare-to-work initiatives. In other words, JOBS' performance requirements offered states incentives and opportunities to choose the best service methods according their own situation. As mentioned by Michalopoulos et al. (2004): "states were required to spend at least 55% of JOBS resources on potential long-term welfare recipients or on members of more disadvantaged groups, including those who had received welfare in 36 of the prior 60 months, those who were custodial parents under age 24 without a high school diploma or GED, those who had little work experience, and those who were within two years of losing eligibility for welfare because their youngest child was 16 or older".

From the welfare recipients' perspective, JOBS was considered a mix between conservative and liberal elements: participation requirements, child care guarantees and investments focused on improving the self-sufficiency of AFDC mothers.

In contrast to previous programs, JOBS emphasized education to a larger extent. The provision of education to any adult on AFDC who lacked a high school diploma or did not demonstrate basic literacy was a requirement. In addition, states were also expected to provide job skills training, job placement services and two of the following: group or individual assistance in locating a job (job search), on-the-job training, or community work experience (workfare). JOBS also introduced the "learnfare" provision, which required teenage custodial parents to participate in educational activities.

## **2.2 National Evaluation of Welfare-to-Work Strategies (NEWWS)<sup>3</sup>**

The National Evaluation of Welfare-to-Work Strategies was a study undertaken by the Department of Health and Human Services. The main objective was to measure the impacts of eleven mandatory welfare-to-work programs that were created following the guidelines imposed by JOBS. These programs were developed in seven locales, each of them having specific characteristics determined by their designers: Atlanta, Georgia; Columbus, Ohio; Detroit and Grand Rapids, Michigan; Oklahoma City, Oklahoma; Portland, Oregon; and Riverside, California. The evaluation measures the impacts of the programs by comparing the outcomes for a treatment group to the outcomes for a control group. Individuals are randomly assigned into the treatment or control group, and those belonging to the treatment group are subject to program requirements. Under JOBS provision, states were free to design the programs as long as they met the provision's requirements. Three particular states (Atlanta, Grand Rapids and Riverside) used two different approaches for the treatment group and then compared the results to the control group. The human capital development approach focused on providing education services, so individuals could get access to better jobs by having a high-school diploma or a degree. The labor force attachment approach prioritized occupational training and quick entry into the labor market, so individuals would gain work experience and use these skills as a stepping stone to better jobs. In one state, Portland, a welfare-to-work program was created mixing services as job search, education and training, and work experience activities.

For all sites, access to activities and requirements are the two main aspects that differentiate treatment group members from control group members. For example, control group members are not eligible for special program services (program being evaluated), but they are eligible for all other employment and training services in the community, as well as for all basic welfare benefits. In the three sites under study, it was observed control group participation in activities as: job search, basic education, college, vocational training and work experience or on-the-job training. With respect to requirements, neither requirements of participation in activities nor sanctions for noncompliance were imposed over control group members. In

---

<sup>3</sup> The discussion in this section is based largely on Freedman et al. (2000), Hamilton et al. (1997), Hamilton et al. (2001) and Scrivener et al. (1998).

contrast, all AFDC recipients part of the treatment group were required to participate if they had access to child care.

In the remaining of this subsection, we will explain the main differences between the activities that treatment and control group members participated in. The three first subsections will focus on the Labor Force Attachment approach, Human Capital Development approach and Portland's employment-focused program, respectively. The last two subsections will explain the participation patterns for treatment and control group members.

### **2.2.1 Labor Force Attachment Approach (LFA)**

The LFA program begins with job search activities, followed by short-term education and training only for those unable to find employment during job search. If necessary, individuals were encouraged to use the first job as a stepping stone in order to get a better work opportunity later.

There were a number of activities considered as part of the LFA approach. As mentioned earlier, each stage had as goal to help individuals to get a job. In case individuals didn't succeed or the time limits of the program run out, participants would be placed into the next stage of the program. Let's review each of these stages:<sup>4</sup>

*Job Club:* this program lasted between three and five weeks, where the two main components were classroom instruction and phone room. In this stage, clients were taught how to find job leads and fill job applications, how to write a resume and a cover letter, how to conduct an interview, and how to value their talents. Classroom instruction lasted between one week (Riverside, including an in-depth comparison of welfare and earned income) and two weeks (Grand Rapids, including career exploration). Clients participated in these classes from 15 to 30 hours per week depending on the site under evaluation. In addition, participants were expected to show up on time to classes and come dressed as they would for a job. The phone room segment came after the classroom instruction and the goal was to improve clients' job-seeking skills by calling potential employers, getting interviews and writing job applications. The sites provided a wide variety of support services in order to help clients to find jobs. For example, telephones were available at each site so that participants could call employers and

---

<sup>4</sup> For more details, see Appendix Figure A.1.

receive messages. Classified advertisement sections from newspapers, telephone directories and job announcements were also available at JOBS offices.

*Individual job search:* this section lasted from three to five weeks per year. Clients were required to look for employment by themselves, write down the names of the companies they had contacted and report their progress each week to the staff. The number of employer contacts required was dependent on individual characteristics and it was determined by program staff. As mentioned earlier, individual search was the second stage in the program and it was provided to individuals who did not find job after completing job club. The length of the two first stages combined, *job club* and *individual job search*, was eight weeks maximum per year.

*Basic education or vocational training:* this stage lasted nine months maximum. Basic education services were provided to clients who did not have a high school diploma or a GED certificate. In Riverside, these services were provided to clients who possessed these credentials but with low scores. Individuals were assigned to four major types of classes based on their achievements: high school completion, General Education Development (GED), Adult Basic Education (ABE) and English as a Second Language (ESL). In Grand Rapids, vocational training was available with short programs generally leading to a certificate of credit.

*Work experience:* it lasted from three to six months and included three types of positions. Firstly, unpaid work was the most common work experience where individuals were assigned to public or private non-profit sectors. Secondly, it was also offered on-the-job training in the private sector with a wage subsidized by the client's welfare grant. Thirdly, paid work was also an option usually in the form of college work-study positions. Unpaid work experience positions were developed by JOBS staff and clients were normally assigned after completing job club or other activities without finding work.

## **2.2.2 Human Capital Development Approach (HCD)**

The HCD program begins with longer-term education and training, generally lasting up to two years. Job search and vocational training activities may be assigned if clients do not find

employment through their education and training program, or on their own initiative. Let's review the stages:<sup>5</sup>

*Basic education:* this stage lasted from six months to one year and it was quite similar to the one considered in the labor force attachment approach. Basic education services were provided to clients who did not possess a high school diploma or a GED certificate. In Riverside, clients who possessed these credentials but with low scores were required to participate in these services. As before, there were four major types of classes: high school completion, General Education Development (GED), Adult Basic Education (ABE) and English as a Second Language (ESL).

*Vocational training or college:* this stage lasted up to two years. Individuals who wanted to participate in training programs or get an academic degree started here. The principal providers of vocational training were public schools and community colleges, among others. In Grand Rapids, vocational training included different training programs as: automotive maintenance and repair, business and clerical occupations, cabinet and furniture making, computer programming, cosmetology, electronics, nursing, refrigerator repair, and truck driving. Most of these programs lasted between one and two years, where one-year degrees led to a certificate of credit and two-year programs led to an undergraduate academic degree. College enrollment was limited to clients who could complete an associate's degree<sup>6</sup> within two years. In Riverside, clients were not allowed to obtain vocational training unless they had enrolled on their own initiative before random assignment.

*Work experience:* this stage lasted from three to six months. Individuals preferring work experience started here. As mentioned earlier, this stage included three types of positions. The first and most common type was unpaid work in the public or private non-profit sectors. The second type was on-the-job training in the private sector where a wage subsidized by the client's welfare grant was usually offered. The third type was paid work experience, normally in the form of college work-study positions.

*Job search:* it lasted up to eight weeks per year. The activities were quite similar to those experienced by the LFA treatment group and it included job club plus individual job search.

---

<sup>5</sup> For more details, see Appendix Figure A.2.

<sup>6</sup> An associate's degree is an undergraduate academic degree lasting normally two years and awarded by community colleges, technical colleges and bachelor's degree-granting colleges and universities.

### **2.2.3 Portland's employment-focused program**

The program implemented in Portland followed an employment-focused approach, having employment as the primary goal. Portland's program emphasized full-time jobs that paid more than the minimum wage, included benefits and offered room for advancement. There was also heavy focus on job development and placement activities. Furthermore, the strategy was to provide a mix of different services to participants: job search, education and training, and work experience activities. In other words, the strategy was to provide a blend of strong LFA elements and moderate HCD elements.

As in other states, program group members attended a group JOBS orientation immediately following skills testing. Clients were selected by managers to attend two different service tracks: fast track or enhanced track. Selection was based on a variety of factors as employment history, educational status and personal goals. Clients that were ready to look for a job were placed in the fast track. In this track, clients participated in activities as job club and job search. The enhanced track included clients that were not ready to enter the labor market. For this reason, clients in the enhanced track participated in life skills training classes and basic education classes instead.

In the rest of this subsection, we will explain briefly the main activities:<sup>7</sup>

- a) Job Club / Job Search: consisted of 30 hours of classes per week during a period of two weeks. Some of the topics discussed were: career goals, resume preparation and videotaped practice interviews.
- b) Individual job search: this activity took place right after job search. Clients were allowed to use resource rooms and were sometimes assigned to a staff member who monitored and assisted them in their search.
- c) Life skills training: this four- to five-weeks class involved examination of work history and vocational interests. The main goal was to prepare people for work and eventual self-sufficiency.
- d) Basic education: basic education services were provided primarily to clients who did not have a high school diploma or GED. For example, a six-week GED class was

---

<sup>7</sup> For more information, see Appendix Table A.1 and Appendix Table A.2.

provided to individuals missing the GED test in social studies, literature, science, mathematics or writing. Clients whose achievements were lower than what is required for high school completion were assigned to Adult Basic Education (ABE).

- e) Work experience: considered activities as unpaid work in the non-profit and private sector, on-the-job training in the private sector and paid work. Participation was voluntary in unpaid positions, and positions lasted a maximum of three months.

## **2.2.4 Participation patterns for treatment group members**

AFDC recipients are exposed continuously to a decision-making process, where welfare payments are contrasted to labor earnings. More specifically, treatment group members can decide between participating in JOBS activities and continue receiving AFDC payments, or not participate in the mandatory activities and receive grant sanctions. As expected, the decision will depend on the gains obtained from each alternative. Since the impacts of a program are commonly measured as the average effect over the sample population, non-compliers will decrease the average effect of a training program. For this reason, it is of interest to understand clearly the mechanism behind the decisions of treatment group participants.

When mapping the decisions of program participants,<sup>8</sup> the first step is to identify all the different options or decision nodes that a treatment group member has. The second step is to identify the gains obtained from following each path. The third step is to check if there are any “dominated” strategies, meaning if there is any path that would never be followed by an individual because it leads to low gains. After clarifying all the paths, you can set incentives or sanctions in order to motivate an individual to follow a determinate path, i.e., exit welfare.

In this subsection, we will explain briefly all the different tracks that can be followed by a treatment group member and then explain in some detail the three main tracks.

For individuals participating in training activities, two main tracks can be followed: employment leading to exit from AFDC and no exit from AFDC.<sup>9</sup> Individuals exiting AFDC receive 12 months of transitional services and are still eligible to receive food stamps.

---

<sup>8</sup> For more details, see Appendix Figure A.3 and Appendix Table A.3.

<sup>9</sup> Exit from AFDC is defined as two consecutive months of zero payments recorded on the state AFDC administrative records system.



Individuals that do not exit AFDC and JOBS participation continues being mandatory, have two main options: to be employed and receive welfare payments or to continue participating in JOBS activities without being employed. As mentioned, the final decision of getting a job or continuing on welfare will depend on the incentives of each option.

Individuals that decided to not participate in activities and that JOBS participation is still mandatory are exposed to sanctions or reductions in welfare payments such as cash, food stamps and Medicaid.

The three main decision paths followed by treatment group participants in Grand Rapids, Portland and Riverside are the following:

*Individuals that exit from AFDC:* Clients moving from welfare to work through exit from AFDC, would stop receiving AFDC benefits but could still be eligible to receive food stamp benefits. In addition, individuals would receive 12 months of transitional assistance such as transitional Medicaid (TMA) and transitional child care (TCC). Explaining in more detail, exit from AFDC is defined as 2 consecutive months of zero payments recorded on the state AFDC administrative records system and occurs when individuals get monthly earnings over \$793 in Riverside and over \$564 in Grand Rapids. States tried also to expand eligibility for transitional Medicaid and child care program through the approval of transitional assistance waivers.

*Individuals who are employed but still AFDC eligible:* Individuals who started working but did not get enough earnings to exit from AFDC could get two different types of benefits: income disregards and food stamps. According the United States Department of Health and Human Services:

“Under the AFDC rules, all recipients who worked were entitled to a \$90 work expense disregard. In addition, for the first four months of AFDC receipt, the next \$30 of earned income, plus one-third of the remainder, was disregarded in calculating eligibility and benefits. After four months and until one year, only the \$30 disregard continued. After one year, there was no earned income disregard. This meant that after one year of AFDC receipt, if a recipient got a job, her grant amount was reduced by one dollar for every dollar that she earned above the amount set aside to cover her work expenses”.

Earned income disregards were created in order to help individuals in their transition to employment. However, the termination of income disregards removed the incentives to work. In other words, without income disregards, welfare payments would be replaced by earnings, keeping total income almost constant. For this reason, many states approved income disregards waivers starting 1992, affecting the whole welfare recipient population in the periods and sites of interest. For example, the state of California (Riverside) adopted flat earnings disregard of \$120 and a percentage earnings disregard of 33.33%, starting in October 1992. The state of Michigan (Grand Rapids) adopted a flat earnings disregard of \$200 plus a percentage earnings disregard of 20%, starting also in October 1992. The state of Oregon (Portland) increased gross income limits to 130% of federal poverty guideline for JOBS Plus participants, starting in July 1995.

In the case of food stamps, grant calculations count a dollar of earnings less than a dollar of AFDC, so a person that replaces welfare dollars with earnings may experience a net increase in food stamps. The food stamp benefit level is calculated by considering the maximum benefit level minus one-third of a household's countable income. The countable income includes 100% of AFDC payments but only 80% of earnings, so an individual who replaces AFDC with earnings could lower her countable income and increase her food stamp payments. On the other hand, it is also possible that a recipient might decrease or completely lose food stamps benefits if earning gains are relatively large.

*Non-compliers receiving sanctions:* Even though states were allowed to design their own programs, federal JOBS regulations governed the enforcement rule and sanctioning process nationwide. As mentioned by Hamilton et al. (1997), the penalty for noncompliance was removal of the JOBS-mandatory client from the AFDC grant. For example, in case that the parent failed to participate, the AFDC grant was reduced so only the children were covered. Sanctions were to continue until the sanctioned individual complied with the participation mandate, with a minimum sanction length of one month for the first "offense", three months for the second, and six months for the third and subsequent offenses. For a three-person family in 1993, a sanction would have resulted in an \$88 decrease in a monthly grant of \$474 in Grand Rapids. In Portland, a sanction would have resulted in a \$142 decrease in a monthly grant of \$460 and in Riverside, a sanction would have resulted in a \$120 decrease in a monthly grant of \$624 (Hamilton et al., 1997; Scrivener et al., 1998).

### **2.2.5 Participation patterns for control group members**

As mentioned, the main differences between treatment and control group members are access to activities and requirements of participation. Control group members could not enter the program being evaluated but could receive any services that existed prior to the introduction of the JOBS program. In addition, neither participation requirements nor sanctions were imposed over control group members. These two main characteristics alter the gains obtained from each decision track. However, we will describe to some extent the three main tracks<sup>10</sup>:

*Individuals that exit from AFDC:* Control group members moving from welfare to work would receive the same services as treatment group members. Individuals would have access to transitional Medicaid and child care services for up to 12 months. In addition, they could still be eligible to receive food stamp benefits, depending on the earnings level.

*Individuals who are employed but still AFDC eligible:* Control group members employed and on AFDC, would receive the same benefits as treatment group members. Income disregards and food stamps would be provided in case of eligibility.

*Individuals who still continue on AFDC:* Since control group members were neither subject to requirements nor to sanctions, individuals could choose to not participate in any activities and still continue on AFDC.

## **2.3 Literature Review**

Most studies have tended to focus on one type of estimators, experimental or non-experimental, while a smaller part of the international research has evaluated non-experimental estimators' performance against randomized results. However in the last decades, this tendency has been changing, emphasizing a comparison between methods instead. In his seminal paper, LaLonde (1986) developed the idea of evaluating econometric procedures, where results obtained using non-randomized data were compared to results arising from random-assignment. This idea is followed in our study.

One of the earliest papers published on the historical progression of the welfare reforms in the United States, measures the impacts using the difference of the average outcomes by treatment status specification (Gueron, 1990). In other words, average outcomes of program

---

<sup>10</sup> For more details, see Appendix Figure A.4.

participants were compared to average outcomes of non-participants. Later on, this method was also applied to the NEWWS data set, obtaining experimental estimators where programs' and sites' performance were compared (Hamilton et al., 1997; Scrivener et al., 1998; Freedman et al., 2000; Hamilton et al., 2001). Normally, impacts of training programs were measured as a change in labor earnings and employment status experienced by the program participant. However, the field of program evaluation has also extended to measuring the impacts of training programs on child outcomes (Hamilton et al., 2001). It has been widely accepted that also the families of welfare recipients are affected by policy reforms.

Other econometric methods, as differences-in-differences estimators, were used when calculating the impacts of a mandatory program in the United Kingdom (Blundell et al., 2004). Moffitt (2002) showed empirically that the impacts of labor market reforms in the United States can have opposite effects. In other words, the mean impacts of a labor market reform may average together positive and negative labor supply responses, possibly obscuring the extent of welfare reform's effects. For example, the TANF program impacts on the bottom of the distribution were significantly different to average impacts, suggesting the potential for distributional estimators (Moffitt, 2008). It was also proved that the program had positive impacts on employment levels but no impact on income levels, given that increased earnings were cancelled out by the loss in welfare benefits.

Inspired by the seminal work of LaLonde (1986), a number of studies have applied randomized experiments to evaluate the empirical models used to estimate the impacts of labor market programs (Michalopoulos et al., 2004; Greenberg et al., 2005). Michalopoulos and colleagues (2004) report short run and medium run biases obtained using propensity-score matching methods and compare their results to Freedman et al. (2000). The analysis uses NEWWS data set and indicates, among other conclusions, that the estimated medium run bias is larger than the short run bias. Using a slightly different data set and meta-analysis instruments, Greenberg et al. (2005) explain why Riverside and Portland's programs performed better than the programs assessed in other sites. They conclude that the superior performance of the programs is only partly attributable to the design of these programs with focus on job search and sanctions. Furthermore, they indicate that caseload characteristics might be more important when achieving success.

Section 3 will review the empirical strategies or methods used to calculate the impacts of the programs in the different sites.

### 3 Empirical Strategy

As mentioned previously, our goal is to estimate the impacts of welfare-to-work programs. Predicting this causal relationship allows foreseeing the effects of different programs or circumstances, so policy makers can make better decisions. For example, the causal effect of a training program is the increment to earnings an individual would receive if she participated in the services associated to this program. The ideal procedure to measure this effect would hold everything constant and only manipulate the treatment status. In other words, we would obtain the difference between potential outcomes by going back in time and changing the person's treatment. Since this procedure is not possible, a range of studies have used randomized experiments as a good alternative to solve the selection problem. Even though randomized trials might be the best alternative to measure impacts, experimental data is not always easy available to researchers. As an alternative to randomization, methods using non-experimental data have been developed to calculate impacts. For example, matching methods as propensity-score matching select on observable characteristics. Other methods as differences-in-differences estimators aim to control for unobservable characteristics.

In order to calculate the impacts of a training program over a sample of individuals we need to: (1) identify the characteristics of the treatment or program, (2) identify the selection rule that assigned individuals to a control or treatment group and, (3) choose an estimator and econometric procedure to calculate the impacts of the program.

Since the main characteristics of the training programs were explained in Section 2, we will explain the different selection rules used to assign individuals and the estimators used.

The best evaluation method consists of randomizing sample members into a control or treatment group. In other words, randomization is the first-best of the selection rules because it provides experimental data. For example, experimental data allow us to calculate causal effects without the presence of selection bias. However, randomized data are not always easy available for researchers. For this reason, another option is to use non-experimental data. A comparison group obtained from non-experimental data tries to mimic the properties of the control group in a randomized experiment.

In the following subsections, we will explain these topics in more detail and follow closely the methodology used by Angrist and Pischke (2009), Blundell and Costa Dias (2009) and Wooldridge (2002).

The question under analysis is whether earnings are affected by program participation. Let's start by describing program participation by a binary random variable,  $D_i = \{0,1\}$ , where  $D_i = 1$  indicates that individual  $i$  participates in the program and  $D_i = 0$  indicates that individual  $i$  does not participate in the program. For any individual, there are two potential outcomes:  $Y_i^1$  if  $D_i = 1$  and  $Y_i^0$  if  $D_i = 0$ . In other words,  $Y_i^0$  represents the earnings of an individual who did not participate in the program, while  $Y_i^1$  represents the earnings of the same individual had she participated in the program. The difference between  $Y_i^1$  and  $Y_i^0$  would represent the causal effect of participating in the program for individual  $i$ . The observed outcome  $Y_i$ , can be written as a combination of potential outcomes:

$$Y_i = Y_i^0 + (Y_i^1 - Y_i^0)D_i. \quad (3.1)$$

This particular notation is useful because it contains the causal effect of program participation, i.e.  $Y_i^1 - Y_i^0$ . Since we never observe both potential outcomes for an individual, we estimate the effects of program enrollment by comparing the observed average earnings of participants versus observed average earnings of non-participants. This term is also known as the *average treatment effect* (ATE). Furthermore, we can express the ATE as the sum of the *average treatment effect on the treated* (ATT) and the *selection bias*,

$$\begin{aligned} E(Y_i | D_i = 1) - E(Y_i | D_i = 0) &= \underbrace{E(Y_i^1 | D_i = 1)}_{\text{Average Treatment Effect}} - \underbrace{E(Y_i^0 | D_i = 1)}_{\text{Average Treatment Effect on the Treated}} \\ &\quad + \underbrace{E(Y_i^0 | D_i = 1) - E(Y_i^0 | D_i = 0)}_{\text{Selection Bias}} \end{aligned} \quad (3.2)$$

The ATT reflects the average difference between the earnings of the participants and their earnings in case they did not participate. The *selection bias* represents the difference in average pre-treatment earnings between those who were and those who were not participants.

As we can see, if we want to identify the difference between potential outcomes or ATT, we would not be able to find it directly in the data because of the presence of selection bias. One method to solve this problem is by random assignment. In this case, ATE is also equal to ATT given that randomization makes  $D_i$  independent of the potential outcomes. This topic will be explained in more detail in the first subsection. In case there is non-random selection, we find *selection on the observables* and *selection on the unobservables*. These two topics will be explained in the second and third subsections respectively. In addition, the selection process is expected to have larger consequences in the presence of heterogeneous effects. In the heterogeneous effect model, the treatment group members and the control group members might benefit differently from program participation, such that the ATT differs from the ATE. Thus, heterogeneity can be another reason (than selection bias) explaining why treatment effects might differ. However, in interpreting the estimates, we will disregard the presence of heterogeneity.

### 3.1 Random Assignment

Let's suppose it is possible to run a social experiment where individuals are randomly assigned to a treatment group or a control group. If random assignment is correctly implemented, then we are able to rule out the bias arising from self-selection. In the case of the implementation of a welfare-to-work program, some individuals might be randomized into the training program, while the rest is excluded from the services provided by the program. Since assignment to treatment is random, then program participation would be independent of the outcome or the program effect.

Explaining this more formally, random assignment to program participation solves the selection bias problem because randomization makes  $D_i$  independent of potential outcomes. To check this, note that the independence of  $Y_i^0$  and  $D_i$  means:

$$E(Y_i^0 | D_i = 1) = E(Y_i^0 | D_i = 0)$$

Simplifying equation (3.2) further to:

$$\begin{aligned} E(Y_i | D_i = 1) - E(Y_i | D_i = 0) &= E(Y_i^1 | D_i = 1) - E(Y_i^0 | D_i = 0) \\ &= E(Y_i^1 - Y_i^0 | D_i = 1) = E(Y_i^1 - Y_i^0), \end{aligned} \quad (3.3)$$

the effect of randomly assigned program participation on the participants is the same as the effect of program participation on a randomly chosen participant. In other words, the average treatment effect is equal to the average treatment effect on the treated when  $D_i$  is randomly assigned:

$$ATE = E(Y_i^1 - Y_i^0) = E(Y_i^1 - Y_i^0 | D_i = 1) = ATT.$$

Causality can also be studied by using regression analysis. While controlling for covariates should not affect the consistency of the estimators under random assignment, it may help reduce noise and therefore make the estimators more precise. For example, let's suppose constant treatment effects and express observed outcomes as a combination of potential outcomes as in equation 3.1. It is straightforward to see that equation (3.1) is equivalent to:

$$Y_i = E(Y_i^0) + (Y_i^1 - Y_i^0)D_i + [Y_i^0 - E(Y_i^0)] \quad (3.4)$$

Thus we can replace the terms in equation (3.4), obtaining:

$$Y_i = \alpha + \rho_i D_i + \eta_i, \quad (3.5)$$

where

$$\begin{aligned} \alpha &= E(Y_i^0), \\ \rho_i &= (Y_i^1 - Y_i^0), \\ \eta_i &= Y_i^0 - E(Y_i^0). \end{aligned}$$

Equation (3.5) is a general model, since no functional form or assumptions have been imposed. In this fashion,  $\alpha$  represents the expected untreated outcomes,  $\rho_i$  reflects the difference between the potential outcomes or causal effect of the program, and  $\eta_i$  represents



the random part of  $Y_i^0$ . If we evaluate the conditional expectation of this equation considering treatment status and control status, then we get:

$$\begin{aligned} E(Y_i | D_i = 1) &= \alpha + \rho_i + E(\eta_i | D_i = 1), \\ E(Y_i | D_i = 0) &= \alpha + E(\eta_i | D_i = 0). \end{aligned}$$

These equations imply that,

$$E(Y_i | D_i = 1) - E(Y_i | D_i = 0) = \rho_i + E(\eta_i | D_i = 1) - E(\eta_i | D_i = 0). \quad (3.6)$$

Thus the treatment effect is represented by  $\rho_i$  and the second and third terms on the RHS of equation (3.6) account for the selection bias. As presented above, selection bias captures the correlation between the regression error term  $\eta_i$  and the regressor  $D_i$ . But since,

$$E(\eta_i | D_i = 1) - E(\eta_i | D_i = 0) = E(Y_i^0 | D_i = 1) - E(Y_i^0 | D_i = 0),$$

selection bias reflects the difference in (no participation) potential outcomes between those who participate in the program and those who do not. For example, in program evaluation, selection bias might arise when participation is voluntary given that volunteers might have intrinsically more difficulties finding a job than non-volunteers. Another example is selection bias arising from mandatory training programs, where individuals are non-randomly assigned to the treatment group based on certain criteria that makes them qualify as disadvantaged (lone mothers, ethnicity, education level, welfare recipients, etc.).

If we implement random assignment in this type of model, we ensure that the treatment group members and the control group members are equal in all aspects except in the treatment status. In order to obtain randomization, two key assumptions must hold:

$$E(\eta_i | D_i = 1) = E(\eta_i | D_i = 0) = E(\eta_i), \quad (A.1)$$

$$E(\rho_i | D_i = 1) = E(\rho_i | D_i = 0) = E(\rho_i). \quad (A.2)$$

Assumption (A.1) is better known as *no selection on untreated outcomes*, where individuals' untreated outcomes are not determined by their treatment status. Assumption (A.2) is better

known as *no selection on the (expected) gains*, in other words, expected gains do not determine participation.

We can also express the terms in assumption (A.2) as:

$$\begin{aligned}\rho^{ATE} &= E(\rho_i), \\ \rho^{ATT} &= E(\rho_i | D_i = 1), \\ \rho^{ATNT} &= E(\rho_i | D_i = 0).\end{aligned}$$

Meaning that randomization eliminates the selection bias in (3.6) because of assumption (A.1) and randomization ensures that we identify the average treatment effect because of assumption (A.2). In this fashion, when calculating the causal effects of training programs using OLS, the estimation of the parameter  $\rho_i$  reflects the average treatment effect ATE.

Until now we have only talked about the first-best procedure of program evaluation, randomized trials. In the following subsection, non-random selection methods as *selection on observables* will be reviewed.

## 3.2 Selection on observables

As explained in the past section, the causal relationship between training programs and earnings tells us what individuals would earn on average if we could change their program participation in a completely controlled environment. If we want to generalize this notion to more complicated situations where control variables must be held fixed for causal inference to be valid, we have to state the *conditional independence assumption* (CIA). The conditional independence assumption is also called selection on observables because the covariates to be held fixed are assumed to be known and observed.

As stated at the beginning of Section 3, we can write the estimator for the causal effects of program participation on earnings as:

$$E(Y_i | D_i = 1) - E(Y_i | D_i = 0) = E(Y_i^1 - Y_i^0 | D_i = 1) + E(Y_i^0 | D_i = 1) - E(Y_i^0 | D_i = 0).$$

If participants are randomized into the treatment, then the selection bias is equal to zero, this means that the pre-treatment outcomes for participants and non-participants are equal.

However, if participants are not randomized into the training program, then selection bias can be positive or negative. For example, let's consider a training program for female AFDC recipients where female recipients were randomized into the treatment. If we measure the causal effects of this program by comparing the participants' earnings to male AFDC recipients' earnings, then the selection bias might be negative. In other words, it could be expected that the pre-treatment earnings of male non-participants were larger than the pre-treatment earnings of female participants. In this case, the ATE estimator would underestimate the benefits of the program. However, if we control for gender (observable characteristic) and compare the earnings of female participants to female non-participants, then the source of selection bias would disappear. This concept is considered in the *conditional independence assumption* or *selection on observables*.

The CIA states that conditional on observed characteristics  $X_i$  (gender, age, education, income), selection bias disappears. Stating this formally,

$$\{Y_i^0, Y_i^1\} \perp D_i \mid X_i, \quad (3.7)$$

which means that given the CIA, conditional-on-  $X_i$  comparisons of average earnings across program participation have a causal interpretation. In other words,

$$E(Y_i \mid X_i, D_i = 1) - E(Y_i \mid X_i, D_i = 0) = E(Y_i^1 - Y_i^0 \mid X_i) \quad (3.8)$$

Now, going back to the constant treatment effect assumption, we can calculate the causal relationship by using a regression like (3.5). However, since equation (3.5) is a causal model, program participation  $D_i$  might be correlated with the error term  $\eta_i$ . Let's suppose now that the CIA holds given a vector of observed characteristics,  $X_i$ . Then decompose the random part of potential earnings  $\eta_i$ , into a linear function of observable covariates  $X_i$  and an error term,  $v_i$ :

$$\eta_i = \gamma X_i + v_i,$$

where  $\gamma$  is a regression coefficient that is assumed to satisfy  $E(\eta_i | X_i) = \gamma X_i$ . Given that  $\gamma$  is defined by the regression of  $\eta_i$  on  $X_i$ , the residual  $\nu_i$  and  $X_i$  are uncorrelated by construction. Therefore, the residual in the linear causal model

$$Y_i = \alpha + \rho_i D_i + \gamma X_i + \nu_i \quad (3.9)$$

is uncorrelated with the regressors,  $D_i$  and  $X_i$ , and the regressor  $\rho_i$  is the causal effect.

Emphasizing once again, the key assumption is that  $\eta_i$  and  $D_i$  are correlated only through the observable characteristics  $X_i$ . This is the well-known CIA or selection-on-observables assumption.

Once we are clear about the linear causal model specification, there are many different econometric procedures that can be used in order to estimate the regressor  $\rho_i$  or causal effect of the treatment. The simplest procedure is *ordinary least squares*, where the OLS estimator for  $\rho_i$  is obtained by minimizing the *sum of squared residuals* (SSR). In other words, this method minimizes the sum of squared differences between the observed outcomes in the dataset and the outcomes predicted by the linear approximation.<sup>11</sup>

Even though OLS is by far the most common procedure to control for selection on observables, it may be fragile to model misspecification. For this reason, we continue our analysis by using matching estimators. Since (non-parametric) matching estimators do not require a specific functional form of the outcome equation, there is in principle no bias arising from misspecification.

In the following subsection, we will explain semi-parametric and non-parametric estimation based on the *propensity-score*. Propensity-score matching (PSM) can be applied over non-experimental data and calculates the causal effects while controlling for observable characteristics. As mentioned, causal interpretation of regression coefficients and matching strategies are based on the conditional independence assumption. Indeed, matching and regression can both be considered as control strategies, where regression is a particular sort of weighted matching estimator.

---

<sup>11</sup> The implementation of ordinary least squares will be explained in detail in Section 5.

### 3.2.1 Semi-parametric and non-parametric estimation

#### Matching and propensity-score matching

Matching is an empirical tool where control participants are compared or matched to treatment participants based on similar observed characteristics. More specifically, matching controls for covariates based on the *conditional independence assumption*.<sup>12</sup>

The matching estimators introduced by Rosenbaum & Rubin (1983) pair each treatment group member with a control group member. Thus, matching estimators are a weighted average of comparisons across individuals or groups, defined by covariates. In case treatment is independent of covariates, then the regression and matching weighting schemes would be equal. Another interesting fact is that neither matching estimators nor regression estimators give any weight to covariate cells that do not contain observations from both the control and treatment group. This condition is called *common support*<sup>13</sup> and implies that estimators are limited to covariate values where both control and treated observations are found.

In this subsection, we will build the propensity score matching estimators starting with the basic assumptions. First, we will explain in detail the common support condition. Second, we will present the propensity score theorem. Third, we will establish identification of the average treatment effects estimator by using inverse propensity score weighting. To finish, the propensity-score matching estimator will be presented.

When evaluating the causal effects of training programs, the starting point is a comparison between the individual's potential outcomes with treatment and without treatment. Because we cannot observe both states, given that the individual is a participant or not, then we can understand the problem as one of missing data. To approach this missing data problem, inverse probability weighting methods will be applied later on.

The following approach assumes that we have an independent, identically distributed sample from the population. In other words, we are not considering cases where the treatment of one unit influences another unit's outcome. This assumption is known as the stable unit treatment value assumption (SUTVA) and it states that the treatment of unit  $i$  affects only the outcome

---

<sup>12</sup> The *conditional independence assumption* is also known as *ignorability*, *unconfoundedness* or *selection on observables*.

<sup>13</sup> The *common support* condition is also known as *overlap* assumption.

of unit  $i$ . In addition, we assume that the *conditional independence assumption* and *common support* condition hold.

The *common support* assumption states that for any setting of the covariates in a certain population, there is a chance of seeing units in both the control and the treatment groups. Formally, for all  $X_i \in X$ , where  $X$  is the support of the covariates,

$$0 < P(D_i = 1 | X_i) < 1. \quad (3.10)$$

For example, if  $P(D_i = 1 | X_i = X_i') = 0$ , then individuals having covariates  $X_i'$  will never be in the treatment group. As a result, we will not be able to calculate the average treatment effects over the population that includes individuals with  $X_i = X_i'$ .

The probability of treatment, as a function of covariates  $X_i$ , plays a central role in the estimation of the average treatment effects. This probability of treatment is usually known as the propensity score, and is defined as:

$$p(X_i) \equiv E[D_i | X_i] = P[D_i = 1 | X_i] \quad (3.11)$$

Given the common support assumption, the propensity score can never be equal to zero or one. As stated by Rosenbaum and Rubin (1983), if we add the CIA, we obtain the propensity score matching theorem:

#### *The Propensity Score Theorem*

*Suppose the CIA holds such that  $\{Y_i^0, Y_i^1\} \perp D_i | X_i$ . Then  $\{Y_i^0, Y_i^1\} \perp D_i | p(X_i)$ .*

In words, the propensity score theorem states that if potential outcomes are independent of treatment status conditional on a multivariate covariate vector  $X_i$ , then potential outcomes are independent of treatment status conditional on a scalar function of covariates, the propensity score,  $p(X_i)$ .

Equivalent to what is stated by the OVB formula<sup>14</sup>, the propensity score theorem states that you need only control for covariates that affect the treatment status. More specifically, you only need to control for the covariate representing the probability of treatment.

Now we have stated some basic assumptions required to estimate causal effects through matching methods. The next step is to identify the average treatment effect estimator by using inverse probability weighting.

Inverse probability weighting is a general approach to non-random sampling applied in the presence of missing data. Briefly explained, it solves the missing data problem when selection probabilities are observed. Stating this more formally, the population mean of any function of the treatment  $D_i$  can be recovered by weighting the selected observation by the inverse of the probability of selection (propensity score).

Let's start by defining two counterfactual conditional means,

$$\mu_0(X_i) = E(Y_i^0 | X_i), \quad \mu_1(X_i) = E(Y_i^1 | X_i) \quad (3.12)$$

In this fashion, we note that the average treatment effect conditional on  $X_i$  can be expressed as:

$$\rho(X_i) = E(Y_i^1 - Y_i^0 | X_i) = \mu_1(X_i) - \mu_0(X_i) \quad (3.13)$$

If we want to establish identification of  $\rho(X_i)$ , we can do this by calculating both counterfactual conditional means in the RHS of equation (3.13). Inverse propensity score weighting allows to calculate the values of  $\mu_0$  and  $\mu_1$ , given the propensity score theorem. According to Imbens (2000), the propensity score theorem implies that instead of having to adjust for all pre-treatment variables, it is sufficient to adjust for the propensity score  $p(X_i)$ .

Noting that  $D_i Y_i = D_i Y_i^1$  and assuming that CIA holds, by iterated expectations,

---

<sup>14</sup> One important finding in regression theory is the omitted variables bias formula, which states that coefficients on included variables are unaffected by the omission of variables, when the variables omitted are uncorrelated with the variables included. This concept was extended by Rosenbaum and Rubin (1983) in their propensity score theorem, where estimation strategies relying on matching use a dummy as the causal variable of interest.

$$\begin{aligned}
E\left[\frac{D_i Y_i}{p(X_i)} \mid X_i\right] &= E\left[\frac{D_i Y_i^1}{p(X_i)} \mid X_i\right] = E\left\{E\left[\frac{D_i Y_i^1}{p(X_i)} \mid X_i, D_i\right] \mid X_i\right\} = E\left\{\frac{D_i E(Y_i^1 \mid X_i, D_i)}{p(X_i)} \mid X_i\right\} \\
&= E\left\{\frac{D_i E(Y_i^1 \mid X_i)}{p(X_i)} \mid X_i\right\} = E\left\{\frac{D_i}{p(X_i)} \mid X_i\right\} \mu_1(X_i) = \mu_1(X_i)
\end{aligned}$$

because  $E(D_i \mid X_i) = p(X_i)$ . In a parallel manner,

$$E\left[\frac{(1-D_i)Y_i}{[1-p(X_i)]} \mid X_i\right] = \mu_0(X_i).$$

Combining these two results, we obtain

$$E\left\{\frac{[D_i - p(X_i)]Y_i}{p(X_i)[1-p(X_i)]} \mid X_i\right\} = \mu_1(X_i) - \mu_0(X_i) = \rho(X_i). \quad (3.14)$$

According Abadie and Imbens (2002), if we average the conditional treatment effect over the marginal distribution of  $X_i$ , we obtain the average treatment effect estimator:

$$\rho^{ATE} = E[\rho(X_i)]. \quad (3.15)$$

Along these lines, identification of  $\rho^{ATE}$  is established by using the propensity score. Since  $D_i$ ,  $Y_i$  and  $X_i$  are all observed, it is straightforward to calculate the causal effects. In practice, however, we need to first calculate the propensity score function  $p(\cdot)$ . Rosenbaum and Rubin (1983) suggest using a flexible logit model to calculate the propensity score, where various interactions of the covariates  $X_i$  are included. For the moment, let  $\hat{p}(X_i)$  denote the estimator of the propensity score, using a logit model. Then, following equation (3.15), we obtain

$$\hat{\rho}_{psw}^{ATE} = N^{-1} \sum_{i=1}^N \left[ \frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1-D_i)Y_i}{1-\hat{p}(X_i)} \right] = N^{-1} \sum_{i=1}^N \frac{[D_i - \hat{p}(X_i)]Y_i}{\hat{p}(X_i)[1-\hat{p}(X_i)]}. \quad (3.16)$$

This estimator is consistent under the conditional independence assumption and the common support assumption.



In the next subsection, econometric methods that control for unobservable characteristics will be explained.

### **3.3 Selection on unobservables**

In randomized experiments, since the assignment to treatment is independent of outcomes, then causal effects can be measured straightforward and no selection bias is found. Assuming that the CIA holds and conditioning on observables, we can still calculate causal effects given that no selection bias is found. However, if we explicitly allow selection to be correlated with unobservables after conditioning on exogenous variables, then we need to use selection-on-unobservables methods. In other words, the CIA fails to hold when individual unobserved characteristics that determine participation, affect the outcome. To address this, we may apply methods developed to control for these unobserved characteristics as time trends, state-specific or individual-specific characteristics, etc.

In the following subsection the differences-in-differences estimator will be explained in detail. Nevertheless, this is not the only econometric procedure that controls for unobservables. There are other popular procedures that control for selection on unobservables: e.g. instrumental variables and fixed effects models. However, these will not be employed in the empirical analysis, and are therefore not discussed here.

#### **3.3.1 Differences-in-Differences (DD)**

When calculating the differences-in-differences estimator, we control for biases that could arise from the comparison between the two groups. As mentioned by Imbens and Wooldridge (2009): “This double differencing removes biases in second period comparisons between the treatment and control group that could be the result from permanent differences between those groups, as well as biases from comparisons over time in the treatment group that could be the result of time trends unrelated to the treatment”.

DD is a version of fixed effects estimation that uses aggregated data by comparing the before and after across groups. The main idea is to use a naturally occurring event to create a shift in the treatment for one group and not another. Thus the difference between these two groups before and after the treatment is compared, creating the DD estimator of the training program.

Normally, DD explores a change in policy occurring at a specific period, which translates in part of the population receiving treatment. Thus longitudinal data are used, where the same individuals are followed over time, before and after the implementation of the program. Since a time dimension is added to the problem, time must be explicitly introduced in the model specification. Each sample member is observed before and after the program implementation, at times  $t = 0$  and  $t = 1$ , respectively. In this context,  $D_{it}$  denotes the treatment status of group member  $i$  at time  $t$  and  $D_i$  (without the time subscript) denotes the treatment group to which individual  $i$  belongs. It is straightforward to notice that  $D_i = 1$  for treatment group members and  $D_{it} = 1$  for treatment group members after program implementation.

Since the DD estimator uses a common trend assumption and assumes no selection on the transitory shock, we can write the outcome equation as:

$$\begin{aligned} Y_{it} &= \alpha + \rho_i D_{it} + u_{it} \\ E[u_{it} | D_i, t] &= E[n_i | D_i] + m_t \end{aligned} \quad (3.17)$$

In equation (3.17),  $n_i$  is an unobservable individual fixed effect and  $m_t$  is an aggregate macro shock. Thus, the DD estimator is based on assumption (A.1) holding in first differences

$$E[u_{i,t=1} - u_{i,t=0} | D_i = 1] = E[u_{i,t=1} - u_{i,t=0} | D_i = 0] = E[u_{i,t=1} - u_{i,t=0}]$$

Equation (3.17) does not exclude the possibility of selection on unobservables, but it restricts its source by ruling out selection based on transitory individual-specific effects. In other words, any possible unobservable individual effects are supposed to be time invariant. In addition, it does not rule out selection on expected gains, i.e. assumption (A.2) does not hold. In this fashion, differences-in-differences estimators only identify the ATT. We will come back to this topic later on.

Following equation (3.17), we can write:

$$E[Y_{it} | D_i, t] = \alpha + E[\rho_i | D_i = 1] + E[n_i | D_i = 1] + m_t,$$

if  $D_i = 1$  and  $t = 1$ . Otherwise,

$$E[Y_{it} | D_i, t] = \alpha + E[n_i | D_i] + m_t.$$

It is straightforward to calculate the DD estimator, by eliminating both  $\alpha$  and the error components through sequential differences

$$\begin{aligned} \rho^{ATT} &= E[\rho_i | D_i = 1] \\ \rho^{ATT} &= \{E[Y_{it} | D_i = 1, t = 1] - E[Y_{it} | D_i = 1, t = 0]\} \\ &\quad - \{E[Y_{it} | D_i = 0, t = 1] - E[Y_{it} | D_i = 0, t = 0]\}. \end{aligned} \quad (3.18)$$

The DD estimator is obtained by subtracting the gain over time in the control group from the gain over time in the treatment group. Double differencing is used to remove the bias associated with a common time trend unrelated to the implementation of the program. In other words, the average difference over time in the control group ( $D_i = 0$ ) is subtracted from the average difference over time in the treatment group ( $D_i = 1$ ).

If we calculate the sample analog of equation (3.18), we obtain the DD estimator:

$$\rho^{DID} = [\bar{Y}_{t=1}^1 - \bar{Y}_{t=0}^1] - [\bar{Y}_{t=1}^0 - \bar{Y}_{t=0}^0] \quad (3.19)$$

where  $\bar{Y}_t^{D_i}$  is the average outcome over group  $D_i$  at time  $t$ . DD measures the excess outcome change for the treatment group members as compared to the control group members, identifying the *average treatment effect on the treated*,

$$E[\rho^{DID}] = \rho^{ATT}.$$

Section 4 describes the data and explains some of the main characteristics of the sample. In addition, it checks balance between control group and treatment group members.

## 4 Data and Descriptive Statistics

This study uses National Evaluation of Welfare-to-Work Strategies (NEWWS) data, which provide wide information for U.S. citizens. The Department of Health and Human Services is the responsible for providing the information for seven locales: Atlanta, Georgia; Columbus, Ohio; Detroit and Grand Rapids, Michigan; Oklahoma City, Oklahoma; Portland, Oregon; and Riverside, California. These sites were nominated by their states and chosen based on four different criteria: (1) prior experience running welfare-to-work programs; (2) an interest in human capital activities; (3) large enough caseloads to provide the sample needed for the experiments; and (4) the commitment to run a random assignment study (Hamilton & Brock, 1994).

There are six main datasets available as part of NEWWS data, however our study focuses on the five-year full impact sample considering three sites: Grand Rapids, Portland and Riverside.

As a general overview, the sample used contains background characteristics as: year of randomization, gender, marital status, number of children, ethnicity, age and education level. Administrative records for each observation are also considered, including 2 years of data prior randomization and 5 years of data after random assignment. Some of the variables included in the administrative records are: earnings, employment status, welfare payments and food stamp payments. Earnings were converted to 1996 dollars using the consumer price index CPI-U (Economic Report of the President, 2000).

Our analysis includes all data for women in Grand Rapids, Portland and Riverside; which had two years of earnings data prior to random assignment. In Riverside, sample members were randomized into control and treatment groups between the second quarter of 1991 and the second quarter of 1993. In Grand Rapids, randomization took place between the third quarter of 1991 and the first quarter of 1994. In Portland, random assignment started in the first quarter of 1993 and ended in the fourth quarter of 1994 (Michalopoulos et al., 2004).

*Comparisons using experimental data:* Control and treatment group members were assigned randomly right after the JOBS orientation, in the case of experimental data. As mentioned earlier, individuals in Portland were randomized into a control or treatment group, where treatment group members participated mainly in employment-focused activities. Individuals

in Grand Rapids were randomly assigned to a control, LFA or HCD group. In Riverside, after obtaining the results from the CASAS exams, individuals with low scores went through a three-way random assignment as in Grand Rapids, while individuals with high scores were assigned to a control or LFA group.

It is of importance to check if random assignment of individuals was successful, i.e., if randomized groups are balanced, with similar pre-treatment characteristics. To do this, we can compare the average pre-treatment values of control and outcome variables among the randomized groups.

Table 1 shows that in Portland, most of the differences between control group and treatment group members are not significantly different from zero. However, there were some exceptions as food stamp payments, number of children and education. Control group members tended to receive higher amount of food stamp payments and more frequently than treatment group members. A higher percentage of program participants had only one child, while a higher share of control group members had three or more children. The treatment group had a higher share of individuals having a high school diploma or GED certificates. Furthermore, when comparing sample sizes, almost as twice as many participants were assigned to the treatment group.

The balancing analysis for the other two sites, Grand Rapids and Riverside, tells the same story. For this reason, these results will only be explained briefly.

Balance was reached for almost all the characteristics in Grand Rapids. Nevertheless, there were some exceptions, where differences were significantly different from zero. For example, control group members received larger AFDC monthly payments than individuals in both treatment groups. A slightly lower percentage of HCD participants had Afro-American ethnicity, when compared to their counterfactuals.<sup>15</sup>

In the case of Riverside, there were no differences significantly different from zero, when comparing groups formed by individuals needing education.<sup>16</sup> For individuals that were not in need of education, control group members earned \$400 extra on average than their

---

<sup>15</sup> For more details, see Appendix Table B.1.

<sup>16</sup> For more details, see Appendix Table B.2.

counterfactuals, during the pre-randomization period. In addition, a higher percentage of treatment group members received AFDC payments the year prior to random assignment.<sup>17</sup>

*Comparisons using non-experimental data:* When applying non-experimental methods, a control group from one site (control group) is compared to a control group from another site (comparison group). As before, balance will be checked by comparing earnings, employment and demographic characteristics among sites.

The top panel of table 2 summarizes earnings and employment characteristics for control group individuals in the three sites. Differences in average annual earnings and average quarters employed in two years prior to random assignment were significantly different from zero for the three comparisons. The largest difference in average earnings prior randomization was \$865, between Riverside and Portland. On average, individuals in Portland were employed one quarter a year, while individuals in Grand Rapids were employed 1,28 quarters a year. Considering years 1 and 2 after randomization, control group members belonging to Portland had significantly higher annual earnings (\$2672) than individuals in Riverside (\$2370). This trend continued through the third to the fifth years, where individuals in Grand Rapids and Portland had average earnings of \$5497 and \$5264, while individuals in Riverside earned \$3875 per year.

The bottom panel of table 2 summarizes the demographic characteristics for control group members that will be used as covariates when calculating regressions. The caseload in Grand Rapids seems to be slightly younger than in Riverside, this might explain the higher share of individuals with one child. Portland has the highest share of white individuals (70%), while Grand Rapids has the higher share of individuals with Afro-American background (41%) and Riverside has the higher share of Hispanics (28%). It might be of interest to notice that the share of Hispanics in Grand Rapids and Portland was set to zero in order to protect the identity of the individuals. In these two sites, individuals with Latin-American background are considered part of the group “not black or white”. Some other differences are observed when reporting marital status and age of the youngest child, however differences between sites might naturally arise given the sample sizes.

Section 5 reviews the implementation of the different econometric methods.

---

<sup>17</sup> For more details, see Appendix Table B.3.

Table 1.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Portland.

Characteristic	Portland		
	Control	Program	Difference
<b>Earnings, welfare payments and employment</b>			
Employed in year prior randomization	0,3981 (0,4896)	0,4139 (0,4926)	0,0159 (0,0143)
Earnings in year prior randomization	1349,3240 (2969,7500)	1429,7810 (3065,0610)	80,4574 (88,3012)
Received AFDC during year prior randomization	0,8502 (0,3570)	0,8488 (0,3583)	-0,0014 (0,0104)
Monthly AFDC received in year prior to RA	375,1936 (189,7974)	369,8660 (188,1109)	-5,3276 (5,4984)
Number of months received AFDC year prior RA	8,1898 (4,7209)	8,0139 (4,7391)	-0,1760 (0,1379)
Received Food Stamps in year prior RA	0,8962 (0,3051)	0,8737 (0,3322)	-0,0224** (0,0094)
Monthly Food Stamps received	219,5522 (102,3303)	211,0973 (106,5724)	-8,4549*** (3,0607)
Number of months receiving Food Stamps prior random assignment	9,1898 (4,3051)	8,8254 (4,5025)	-0,3645*** (0,1291)
<b>Demographic Characteristics</b>			
Single parent, ever married	0,5044 (0,5001)	0,5141 (0,4999)	0,0097 (0,0146)
One child	0,3699 (0,4829)	0,4025 (0,4905)	0,0326** (0,0143)
Two children	0,3328 (0,4713)	0,3363 (0,4725)	0,0035 (0,0138)
Three or more children	0,2973 (0,4572)	0,2612 (0,4394)	-0,0361*** (0,0130)
Any child 0-5 years old	0,7043 (0,4565)	0,6980 (0,4592)	-0,0063 (0,0134)
Black	0,1974 (0,3981)	0,1977 (0,3983)	0,0003 (0,0116)
Not black or white	0,1028 (0,3037)	0,1072 (0,3094)	0,0044 (0,0090)
Age at random assignment	30,7491 (6,4545)	30,5128 (6,5420)	-0,2363 (0,1897)
High school diploma or GED	0,6339 (0,4819)	0,6668 (0,4714)	0,0328** (0,0139)
<b>Sample size</b>	1849	3247	

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

Table 2.- Selected characteristics of female control group members with two years of earnings data prior to random assignment by site.

Characteristic	Riverside	Grand Rapids	Portland	Port-Grand	River-Grand	River-Port
<b>Earnings and Employment</b>						
Average annual earnings in two years prior to random assignment (1996 \$)	2683,80 (5049,72)	2134,67 (4041,11)	1819,05 (3492,95)	-315,62** (132,70)	549,13*** (154,48)	864,75*** (133,84)
Average annual quarters employed in two years prior to random assignment	1,08 (2,68)	1,28 (2,59)	1,00 (2,48)	-0,28*** (0,09)	-0,21*** (0,09)	0,07* (0,08)
Average annual earnings in years 1 and 2 after random assignment (1996 \$)	2369,61 (5030,60)	2518,55 (4311,95)	2672,16 (4402,61)	153,61 (154,92)	-148,94 (156,49)	-302,55** (142,25)
Average annual earnings in years 3, 4, and 5 after random assignment (1996 \$)	3874,76 (6794,84)	5497,09 (6720,90)	5264,28 (6657,16)	-232,82 (237,30)	-1622,33*** (220,17)	-1389,52*** (199,86)
<b>Demographic Characteristics</b>						
Age (years)	32,11	29,65	30,75	1,09	2,45	1,36
Ethnicity						
White	0,51	0,48	0,70	0,22	0,03	-0,19
Black	0,17	0,41	0,20	-0,21	-0,24	-0,03
Hispanic	0,28	0,00	0,00	0,00	0,28	0,28
Not black or white	0,32	0,11	0,10	-0,01	0,21	0,22
Received high school diploma or GED	0,64	0,59	0,63	0,04	0,04	0,01
Single parent, ever married	0,66	0,42	0,50	0,08	0,24	0,16
Number of children						
One child	0,38	0,46	0,37	-0,09	-0,07	0,02
Two children	0,32	0,36	0,33	-0,03	-0,04	-0,01
Three or more children	0,28	0,19	0,29	0,11	0,10	-0,01
Any child 0-5 years old	0,57	0,69	0,70	0,01	-0,12	-0,12
<b>Sample size</b>	2960	1390	1849			

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.



# 5 Implementation

In this section, the implementation of different econometric methods using experimental and non-experimental data will be explained.

Referencing Section 3, experimental data are obtained by randomizing individuals into a control or treatment group, i.e., program participation is independent from the outcomes. Non-experimental data reflect the presence of a selection rule, e.g., selection on observable characteristics or selection on unobservable characteristics. Methods as OLS and PSM rule out selection by controlling for observable characteristics, obtaining unbiased estimators of causal effects. Differences-in-differences rules out selection on unobservables by restricting its source and removing the common trend bias by double-differencing.

When considering the impact of programs over randomized data, we simply compare the outcome variables for control group members to treatment group members. This procedure is straightforward for Portland and Grand Rapids, since there is only one control group. In Riverside, there are 2 control groups and 3 treatment groups in total, so the impacts must be measured using the right comparison.

In the case of non-experimental comparisons, we use control group members from different sites. For example, the control group members in Portland are compared to the control group members in Grand Rapids. This allows calculating the selection bias arising from non-experimental methods as propensity-score matching.

In the following subsections, we will explain in detail the implementation of these procedures.

## 5.1 Random assignment

As mentioned, since individuals in our sample have been randomized into a control and treatment group, causal effects can be measured directly by calculating the difference in mean outcomes between control and treatment group members. Strictly speaking, in the presence of random assignment, the average treatment effect is equal to the average treatment effect on the treated, given that selection bias is equal zero. However, randomized experiments do not always lead to balanced subjects' characteristics across the control and treatment groups.

Thus it is important to check first whether or not randomization was successful. Following this guideline, Section 4 showed that differences across control and treatment groups were small and not significantly different from zero. As a result, we can assume that random assignment worked as expected and proceed to estimate the causal effects of program participation.

In Portland, treatment group members were compared to control group members. In Grand Rapids, there were two comparisons: (1) LFA group members were compared to control group members, and (2) HCD group members were compared to control group members. In Riverside, causal effects for individuals in need and not in need of education were calculated separately. For individuals in need of education, there were two comparisons: (1) LFA participants were compared to control group members, and (2) HCD participants were compared to control group members. For individuals that were not in need of education, there was only one comparison: LFA participants were compared to control group members.

The following regression including control and treatment group members for each site is estimated through OLS:

$$Y_i = \alpha + \lambda P_i + \varepsilon_i, \quad (5.1)$$

where

$Y_i$  = earnings after random assignment for sample member  $i$ ,

$P_i = 1$  if the sample member is in the treatment group and 0 if she is in the control group.

Since randomly assigned program participants are compared to randomly assigned control group members, the difference in average outcomes between the two groups should reflect the average treatment effect of program participation. Thus, the parameter  $\lambda$  in the equation (5.1) provides an estimate of causal effects of program participation. In case the estimated coefficient  $\lambda$  is equal to zero, this would imply that there is no difference in outcomes arising from program participation. In other words, it would mean that the training program did not have any effects on the earnings of its participants.

## 5.2 Ordinary Least Squares (OLS)

In this subsection, ordinary least squares regressions are calculated using non-experimental data. Control group members from one site, were compared to control group members from another site. For example, control group members from Portland were part of the “control group”, while control group members from Grand Rapids were part of the “comparison group”. Following this procedure, we obtain three different comparisons: Portland – Grand Rapids, Riverside – Grand Rapids and Riverside – Portland.

The following regression including the full sample (control and comparison group members) is estimated through OLS:

$$Y_i = \alpha + \lambda C_i + \sum_j \beta_j Z_{ij} + \sum_j \gamma_j W_{ij} + \sum_m \delta_m X_{im} + \varepsilon_i, \quad (5.2)$$

where

$Y_i$  = earnings after random assignment for sample member  $i$ ,

$C_i = 1$  if the sample member is in the control group and 0 if she is in the comparison group,

$Z_{ij}$  = earnings in the  $j$ -th quarter prior to random assignment for sample member  $i$ ,

$W_{ij} = 1$  if the sample member was employed in the  $j$ -th quarter before random assignment, and 0 otherwise,

$X_{im}$  = the  $m$ -th background characteristic for sample member  $i$ .

Since control group members are compared to a non-experimentally chosen control group, there is no average treatment effect on the treated because no individuals are treated. Thus, the parameter  $\lambda$  in the equation above provides an estimate of the mean selection bias for the full sample. In case the estimated coefficient  $\lambda$  is equal to zero, this would imply that there is no selection bias arising from non-experimental comparisons. Thus the model (5.2) would be successful when controlling for observables.

### **5.3 Propensity score matching (PSM)**

In our study, we use propensity score sub-classification methods to estimate the bias resulting from non-experimental mismatch. Since control group members from one site are compared to control group members from another site, no individuals are being treated. This implies that the observed difference in average earnings would represent the selection bias arising from empirical mismatch.

We will calculate the out-of-state bias using the sub-classification approach of propensity-score matching. There were two biases calculated for each comparison: short-run bias covering the two years following random assignment and medium-run bias covering the third through fifth years following random assignment. These two biases are assessed in order to check if non-experimental methods perform better in the near future. This might be of relevance for policy makers because while the impacts of some policies might be evaluated on short-run basis, other programs (like Human Capital Development approach) might be best evaluated on longer-term effects.

In the following, we will use the approach stated by Michalopoulos et al. (2004). The sub-classification approach of propensity-score matching method groups all sample members into subclasses with similar propensity scores. We obtain an estimate of the bias for each subclass from the difference between control group and comparison group outcomes. To estimate the bias for the full sample, we calculate a program participation regression, where dummies for all subclasses except one are included.

First, we estimate a logistic regression of the factors predicting membership in the control group from the full sample. The estimators obtained were then used to obtain a propensity score for each sample member, based on her individual characteristics.

Second, we create subclasses of sample members based on similar propensity scores. This step created five subclasses based on the quintile distribution of estimated control group propensity scores. As a model specification test, comparison group members whose propensity scores were outside the range were eliminated. All control group members remained in the analysis. The rest of the sample members were placed in their respective subclass.

Third, we run a specification test to check if the background characteristics of control and comparison group members are balanced (matched) in each subclass:

$$C_i = \alpha + \sum_j \beta_j Z_{ij} + \sum_j \gamma_j W_{ij} + \sum_m \delta_m X_{im} + \varepsilon_i, \quad (5.3)$$

where

$C_i = 1$  if the sample member is in the control group and 0 if she is in the comparison group,

$Z_{ij}$  = earnings in the  $j$ -th quarter prior to random assignment for sample member  $i$ ,

$W_{ij} = 1$  if the sample member was employed in the  $j$ -th quarter before random assignment, and 0 otherwise,

$X_{im}$  = the  $m$ -th background characteristic for sample member  $i$ .

As we can see, equation (5.3) is analogous to equation (5.2) used when calculating the OLS estimator.

The next step was to conduct an F-test of the joint null hypothesis that all estimated parameters for the equation above are zero except for the intercept. The null hypothesis would then imply that the mean values for all variables in the model are the same (balanced) for both control and comparison groups.

If a subclass was not balanced (the parameters in the model were jointly significantly different from zero at the ten percent level), then it was split in two and tested for balance again. Unbalanced classes were divided until (1) all subclasses were balanced, or (2) further subdividing would result in a subclass with fewer than ten control or comparison group members.

In our study, considering the three particular sites used, we did not have any problems reaching balance. But as mentioned by Michalopoulos et al (2004), in case subclasses remained unbalanced, the process was started again by re-estimating the logistic regression on

the full sample after adding higher order terms, interactions or both. Large t-statistics on coefficient estimates for the unbalanced subclasses were considered when choosing which terms to add. After estimating the new logistic regression, the whole process was repeated over and over again until balance was achieved. If balance could not be achieved, then there was no calculated bias from that comparison.

When balance was achieved, the selection bias for subclass  $k$  was estimated following:

$$Y_i = \alpha_k + \lambda_k C_i + \sum_j \beta_{jk} Z_{ij} + \sum_j \gamma_{jk} W_{ij} + \sum_m \delta_{mk} X_{im} + \varepsilon_i, \quad (5.4)$$

where  $Y_i$  represents earnings after random assignment for individual  $i$ . Even though individuals from the same subclass should have similar values of covariates, the regression adjusts for any small differences that could still exist between the two groups. The parameter  $\lambda_k$  in the equation above estimates the bias in subclass  $k$ . We need to assume that  $\varepsilon_i$  is mean independent of each variable included on the RHS of equation (5.4), in order to obtain a consistent estimate of  $\lambda_k$ . This means that there are no unobserved factors that affect the follow-up earnings  $Y_i$ , earnings before randomization  $Z_{ij}$  and employment levels  $W_{ij}$ . The term  $X_{im}$  includes all the covariates used in the initial logistic regression plus an age-squared term.

As final step, the mean bias for the full sample is estimated by running the following regression:

$$Y_i = \alpha + \lambda C_i + \sum_k^{K-1} \delta_k S_{ik} + \varepsilon_i, \quad (5.5)$$

where

$Y_i$  = earnings after random assignment for sample member  $i$ ,

$C_i = 1$  if the sample member is in the control group and 0 if she is in the comparison group,

$S_{ik} = 1$  if individual  $i$  is member of subclass  $k$  and 0 otherwise.

The parameter  $\lambda$  in the equation (5.5) estimates the average bias for the comparison between the two sites.

## 5.4 Differences-in-Differences (DD)

Differences-in-differences is one of the econometric procedures that controls for unobservable characteristics when using non-experimental data. In our study, regression analysis is used to estimate the causal effects through double differencing by comparing control group members from one site to control group members from another site. As in subsection 5.2, control group members from Portland were part of the “control group”, while control group from Grand Rapids were part of the “comparison group”. Following this procedure, we obtain three different out-of-state comparisons: Portland – Grand Rapids, Riverside – Grand Rapids and Riverside – Portland.

In order to obtain the causal effects through double differencing, we can calculate the following the regression:

$$Y_{igt} = \alpha + \gamma D_{ig} + \lambda D_{it} + \delta(D_{ig} \cdot D_{it}) + \sum_j \rho_j W_{ij} + \sum_m \sigma_m X_{im} + \varepsilon_{igt}, \quad (5.6)$$

where

$Y_{igt}$  = earnings for sample member  $i$ , member of group  $g$ , in period  $t$ ,

$D_{ig} = 1$  if the individual  $i$  belongs to the control group, and 0 if she belongs to the

comparison group,

$D_{it} = 1$  if the observations for individual  $i$  were obtained after program

implementation, and 0 otherwise,

Equation (5.6) includes a group effect represented by  $\gamma$ , a time effect represented by  $\lambda$  and an interaction term  $\delta$  representing the causal effect of the training program. As explained in subsection (3.3.1), by double differencing we obtain the causal effect parameter  $\delta$ .

## 6 Results

There are three key empirical conclusions from this study. First, results obtained using non-experimental data can lead to wrong conclusions about the causal effects of a training program. Second, biases obtained from non-experimental data depend not only on the econometric procedure used but also on the chosen comparison group. Third, comparisons longer ahead in time are more susceptible to selection bias problems. In other words, medium-run bias was larger than short-run bias.

### 6.1 Experimental results

Table 3 presents the impacts of welfare-to-work programs in the three sites of interest. As mentioned earlier, since the conditional independence assumption holds when randomized data is used, impacts can be measured by calculating the difference of means between treatment and control group members. In other words, an unbiased estimator of the causal effect of the program on earnings can be obtained through difference of means given random assignment.

Impacts obtained in the short run were normally higher than those obtained in the medium run and were also significantly different from zero (table 3). In Portland, results in the short run were slightly higher than in the medium run on average. Figure 1 reflects that in pre-randomization quarters, treatment group members' earnings were followed closely by the earnings of control group members. However, program implementation increased significantly quarterly earnings of treated individuals, just to be caught again by control individuals in the last year of the evaluation.

In Grand Rapids, LFA program participants got slightly higher earnings than HCD participants in the short run (table 3). However, medium run differences were not significantly different from zero for none of the programs. In other words, differences between program and control group members tended to dissipate in the last year of the evaluation (figure 2).



Table 3.- Results obtained using experimental data. Impacts are measured through the difference of means between the randomized treatment group and the randomized control group.

		Experimental Data	
		Difference of Means	
		(1)	(2)
Control and treatment groups		Short Run	Medium Run
Portland		1190,88*** (144,98)	1055,63*** (213,81)
Grand Rapids	LFA	466,63*** (158,31)	80,65 (253,58)
	HCD	431,41** (178,22)	228,10 (264,91)
Riverside			
In need of education	LFA	758,37*** (156,06)	453,85** (200,34)
	HCD	323,92** (146,81)	317,03 (195,31)
Not in need of education	LFA	863,51*** (212,90)	230,82 (283,76)

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Note: Short run is defined as the two years following random assignment and medium run is defined as the third through fifth years following random assignment.

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

In Riverside, participants were divided in two groups, depending on the scores obtained in the CASAS exams. For both groups, LFA program showed to be more effective than HCD program when increasing earnings of its participants (table 3). As expected, earnings of individuals not in need of education were always higher than their counterfactuals needing education (figure 3).

Figure 1.- Mean quarterly earnings: Portland

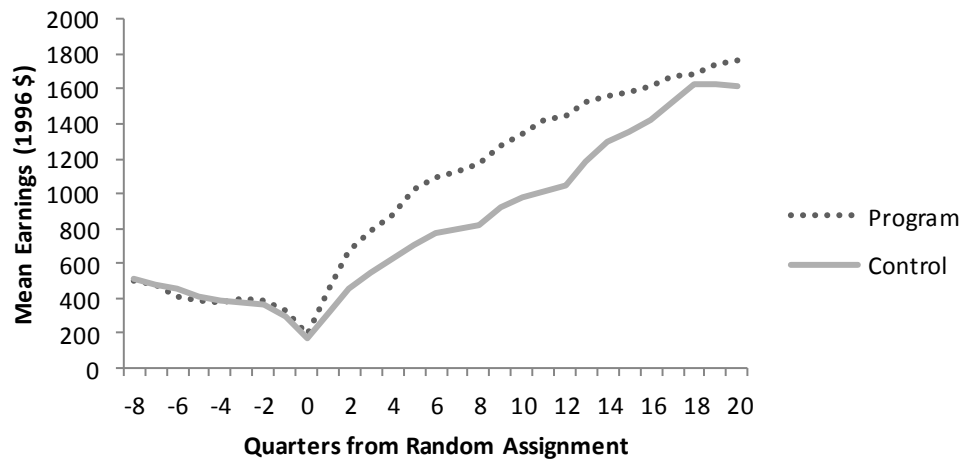


Figure 2.- Mean quarterly earnings: Grand Rapids

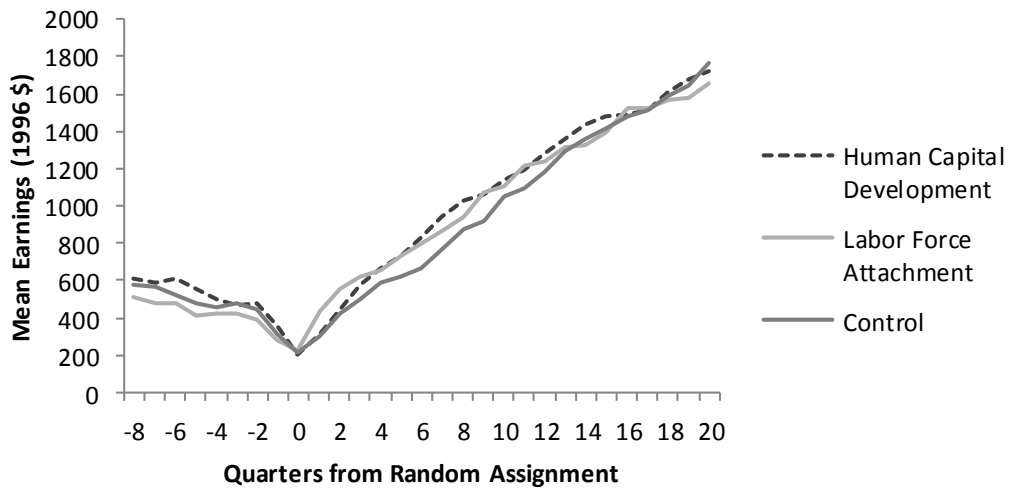
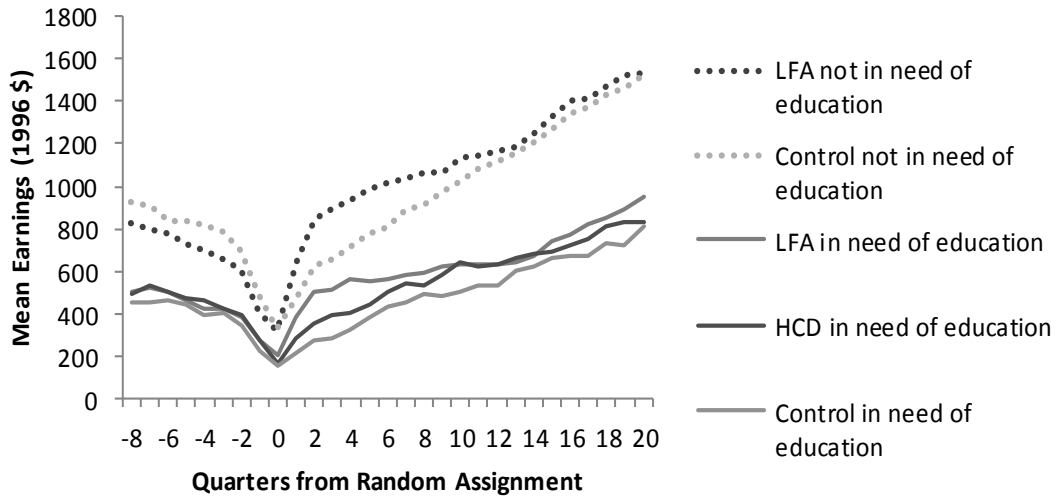


Figure 3.- Mean quarterly earnings: Riverside



## 6.2 Non-experimental results

Results obtained from randomized data are commonly considered a benchmark when comparing results obtained using empirical methods. As explained in Section 3, if we assume that we can control on observable characteristics and the CIA holds, then we can obtain unbiased estimators of causal effects even though participants were not randomly assigned into the control and treatment groups. Another option is to assume the presence of unobservable characteristics and therefore apply methods that calculate the impacts by eliminating the fixed effects.

In this subsection, we estimate the impacts of the program on each site when using non-experimental data. In order to do this, we consider the results obtained from experimental data and the estimated biases obtained from non-experimental data.<sup>18</sup> Causal effects using experimental data were calculated through the difference between mean earnings of a randomized treatment group and mean earnings of a randomized control group. Estimated biases using non-experimental data arose from the comparison in outcomes between a randomized control group and an experimentally chosen comparison group. Thus we can calculate the impacts of the program when using non-experimental data by adding experimental results and non-experimental biases. This would be the equivalent to comparing the outcomes of a randomized treatment group to a non-randomly chosen comparison group.

<sup>18</sup> For more details on the estimated biases using non-experimental data, see subsection 6.3.

The first row of table 4 reflects the experimental results obtained for Portland. Causal effects using experimental data were only calculated using one method: difference of means.

However, causal effects should not vary significantly when calculated using other methods, given that data are randomized. In other words, in an experimental evaluation, estimates of the program effect should not depend on the way earnings and participation equations are specified (LaLonde, 1986). As LaLonde (1986) mentions in his seminal paper: “If the econometric model is specified correctly, the non-experimental estimates should be the same (within sampling error) as the training effect generated from the experimental data, but if there is a significant difference between the non-experimental and the experimental estimates, the econometric model is misspecified”. For this reason, it is of interest to compare experimental and non-experimental estimates.

Table 4 presents estimates for three comparisons in Portland. The first row compares a randomized treatment group to a randomized control group, where both groups are from Portland. The second (third) row compares a randomized treatment group from Portland to an experimentally chosen control group from Riverside (Grand Rapids). Results from the comparison between Portland and Grand Rapids are the closest to experimentally obtained results. However, we can observe that as a general overview, short run and medium run impacts vary widely depending on the comparison site and econometric procedure chosen. Short run impacts vary from \$1348 to \$2327, while medium run impacts vary from \$772 to \$3278. Non-experimental estimations can lead us to results that are as large as three times the experimental results.

The non-experimental results for the other two sites, Grand Rapids and Riverside, tell a similar story. For this reason, the results will only be explained briefly.

In Grand Rapids, the labor force attachment program had impacts ranging from \$16 and up to \$1152 in the short run, and from \$24 to \$2247 in the medium run. Human capital development program’s effects ranged from -\$19 and up to \$1117 in the short run, and from \$172 to \$2394 in the medium run.<sup>19</sup>

In Riverside, there were two main trends arising from non-experimental comparisons. First, in the short run, labor force attachment programs had larger impacts than human capital development programs. Second, most of the non-experimental comparisons showed that both

---

<sup>19</sup> For more details, see Appendix Table C.1.

Table 4.- Comparison between short-run and medium-run impacts measured using experimental and non-experimental data: Portland

Treatment and control group	Selection on observables						Selection on unobservables	
	Difference of Means		OLS Regression		Propensity-score subclassification		Differences-in-differences	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Short Run	Medium Run	Short Run	Medium Run	Short Run	Medium Run	Short Run	Medium Run
Portland (control)	1190,88***	1055,63***						
Portland (Riverside)	1455,16	2397,68	1777,35	2718,51	1686,22	2588,12	2327,32	3278,15
Portland (Grand Rapids)	1348,39	851,09	1423,88	772,18	1419,52	808,65	1641,54	1111,88

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Note: Short run is defined as the two years following random assignment and medium run is defined as the third through fifth years following random assignment  
Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

programs had negative impacts in the medium run. These results might arise from the fact that bias tended to be larger when further away in time. In other words, larger biases were found when comparing control groups from different sites in the medium run. Strictly speaking, the LFA program had impacts ranging from \$73 and up to \$615 in the short run, and from -\$888 to -\$1769 in the medium run, for individuals in need of education. HCD program had impacts from -\$813 to \$180 in the short run, and from -\$1025 to -\$1905 in the medium run. LFA program for individuals not in need of education had impacts from -\$273 to \$720 in the short run, and from -\$1111 to -\$1992 in the medium run.<sup>20</sup>

### **6.3 Estimated bias arising from non-experimental data**

In this subsection, short-run and medium-run biases arising from different econometric specifications are calculated by controlling for observable and unobservable characteristics. The calculations from this section were used to obtain the non-experimental results explained in subsection 6.2.

Let's first check the evolution of mean quarterly earnings for control group members in Grand Rapids, Portland and Riverside. Figure 4 reflects that Riverside had the highest earnings of the three sites before random assignment. However, after the randomization, earnings were higher for control group members in Grand Rapids and Portland. As shown in the figure, earnings of individuals in Grand Rapids were closely followed by the earnings of individuals in Portland for the whole period. This could suggest that a lower bias would arise from the comparison between Grand Rapids and Portland, than the bias arising from the comparison between any other sites.

Table 5 presents the estimated short-run and medium-run bias from different econometric specifications. As a general conclusion, for all sites except one, medium-run bias was larger than short-run bias. In other words, comparisons longer ahead in time were more susceptible to selection bias problems. It can also be observed that for almost all methods, biases calculated between Portland and Grand Rapids were small and not significantly different from zero. This conclusion fits the prediction obtained from the analysis of figure 4.

---

<sup>20</sup> For more details, see Appendix Table C.2.

When comparing Riverside to Portland, all biases calculated were significantly different from zero and varied between -\$264 and -\$2223. Furthermore, calculated biases were smaller when using methods that select on observables. The two methods that minimized short-run and medium-run biases were difference of means and propensity-score subclassification.

The comparison between Riverside and Grand Rapids originated biases starting at -\$143 and up to -\$2166. Again, methods that select on observable characteristics showed to originate lower biases. Furthermore, differences in means and propensity-score subclassification showed to be best econometric procedures.

Biases arising from the comparison between Portland and Grand Rapids were small and not significantly different from zero as a general rule. While methods selecting on observables predicted negative medium-run biases, differences-in-differences predicted a positive bias. However, this result is not especially important given that the biases are not significantly different from zero.

The implications of these results together with the concluding remarks will be discussed in the next section.

Figure 4.- Mean quarterly earnings of control group members in Grand Rapids, Portland and Riverside.

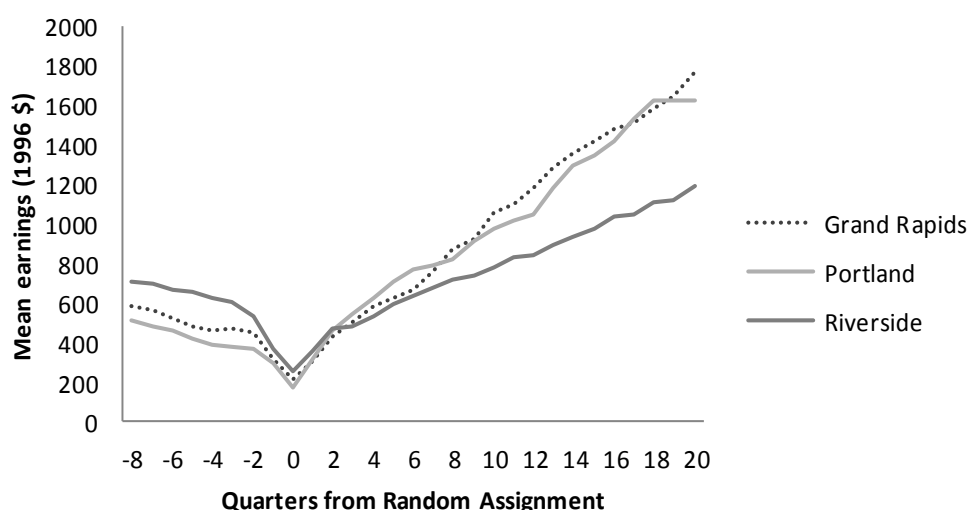


Table 5.- Estimated short-run and medium-run bias for comparisons between control group members in Grand Rapids, Portland and Riverside

Non-experimental Data								
Control and Comparison Site	Selection on observables						Selection on unobservables	
	Difference of Means		OLS Regression		Propensity-score subclassification		Differences-in-differences	
	(1) Short Run	(2) Medium Run	(3) Short Run	(4) Medium Run	(5) Short Run	(6) Medium Run	(7) Short Run	(8) Medium Run
Riverside:								
Portland	-264,28* (143,42)	-1342,05*** (201,17)	-586,47*** (129,90)	-1662,88*** (191,23)	-495,34*** (146,53)	-1532,49*** (206,03)	-1136,44*** (163,35)	-2222,52*** (211,54)
Grand Rapids	-143,44 (158,09)	-1626,52*** (222,27)	-473,19*** (146,87)	-2005,65*** (218,42)	-281,07 (171,44)	-1779,18*** (240,82)	-685,77*** (184,13)	-2166,28*** (233,89)
Portland:								
Grand Rapids	157,51 (151,07)	-240,54 (235,05)	233,00 (145,13)	-283,45 (232,51)	228,64 (159,52)	-246,98 (248,26)	450,66* (174,71)	56,25 (244,57)

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Note: Short run is defined as the two years following random assignment and medium run is defined as the third through fifth years following random. assignment  
Standard errors are shown in parentheses.



## 7 Concluding Remarks

In this study, we estimated the causal effects of a training program over the earnings of control and treatment group members using different econometric procedures. As a benchmark, randomized data were used in order to obtain average treatment effects estimators. Then non-experimental comparison groups were constructed in order to apply methods selecting on observables and unobservables. Thus, causal effects of training programs were calculated using both randomized and non-randomized data.

We might summarize the insights from our research with three broad conclusions. First, results obtained using non-experimental data can lead to wrong conclusions about the causal effects of a training program. Second, biases obtained from non-experimental data depend not only on the econometric procedure used but also on the chosen comparison group. Third, comparisons longer ahead in time are more susceptible to selection bias problems. In other words, medium-run bias was larger than short-run bias.

# References

Abadie, A. and G. W. Imbens (2002): “Simple and bias-corrected matching estimators for average treatment effects”, National Bureau of Economic Research, Technical Working Paper 283.

Angrist, J. D. and J. Pischke (2009): “Mostly Harmless Econometrics: an empiricist’s companion”, Princeton: Princeton University Press.

Blundell, R., M. Costa Dias, C. Meghir and J. Van Reenen (2004): “Evaluating the Employment Impact of a Mandatory Job Search Program”, Journal of the European Economic Association, 2(4), 596-606.

Blundell, R. and M. Costa Dias (2009): “Alternative Approaches to Evaluation in Empirical Microeconomics”, The Journal of Human Resources, 44(3), 565-640.

Economic Report of the President (Washington: U.S. Government Printing Office, 2000).

Freedman, S., D. Friedlander, G. Hamilton, J. Rock, M. Mitchell, J. Nudelman, A. Schweder and L. Storto (2000): “National Evaluation of Welfare-to-Work Strategies: Evaluating Alternative Welfare-to-Work Approaches: Two-Year Impacts for Eleven Programs”, U.S. Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, Manpower Demonstration Research Corporation.

Greenberg, D., K. Ashworth, A. Cebulla and R. Walker (2005): “When Welfare-to-Work Programs Seem to Work Well: Explaining Why Riverside and Portland Shine so Brightly”, Industrial and Labour Relations Review, 59(1), 34-50.

Gueron, J. M. (1990): “Work and Welfare: Lessons on Employment Programs”, The Journal of Economic Perspectives, 4(1), 79-98.

Gueron, J. M. (1991): “From Welfare to Work”, New York: Russell Sage Foundation.

Hamilton, G. and T. Brock (1994): “The JOBS Evaluation: Early Lessons from Seven Sites”, U.S. Department of Health and Human Services, Administration for Children and Families,

Office of the Assistant Secretary for Planning and Evaluation, and U.S. Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education.

Hamilton, G., T. Brock, M. Farrell, D. Friedlander, K. Harknett, J. A. Hunter-Manns, J. Walter and J. Weissman (1997): “National Evaluation of Welfare-to-Work Strategies: Evaluating Two Welfare-to-Work Program Approaches: Two-Year Findings on the Labor Force Attachment and Human Capital Development Programs in Three Sites”, U.S. Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, Manpower Demonstration Research Corporation.

Hamilton, G., S. Freedman, L. Gennetian, C. Michalopoulos, J. Walter, D. Adams-Ciardullo, A. Gassman-Pines, S. McGroder, M. Zaslow, J. Brooks, S. Ahluwalia, E. Small and B. Ricchetti (2001): “National Evaluation of Welfare-to-Work Strategies: How effective are different welfare-to-work approaches? Five-year adult and child impacts for eleven programs”, U.S. Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, Manpower Demonstration Research Corporation.

Imbens, G. (2000): “The Role of the Propensity Score in Estimating Dose-Response Functions”, *Biometrika*, Vol. 87, No. 3, pp. 706-710.

Imbens, G. and J. Wooldridge (2009): “Recent Developments in the Econometrics of Program Evaluation”, *Journal of Economic Literature*, 47(1), 5-86.

LaLonde, R. J. (1986): “Evaluating the Econometric Evaluations of Training Programs with Experimental Data”, *The American Economic Review*, 76(4), 604-620.

Michalopoulos, C., H. S. Bloom and C. J. Hill (2004): “Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?”, *Review of Economics and Statistics*, 86(1), 156-179.

Moffitt, R. (2008): “Welfare reform: the US experience”, Institute for Labor Market Policy Evaluation, Working paper 2008: 13.

Moffitt, R. (2002): “Welfare Programs and Labor Supply”, National Bureau of Economic Research, Working paper 9168.

Office of Human Services Policy (1998): “Aid to Families with Dependent Children: The Baseline”. Available at: <http://aspe.hhs.gov/hsp/afdc/afdcbase98.htm>

Rosenbaum, P. and D. Rubin (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika*, 70, 41-55.

Scrivener, S., G. Hamilton, M. Farrell, S. Freedman, D. Friedlander, M. Mitchell, J. Nudelman and C. Schwartz (1998): “National Evaluation of Welfare-to-Work Strategies: Implementation, Participation Patterns, Costs, and Two-Year Impacts of the Portland (Oregon) Welfare-to-Work Program”, U.S. Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, Manpower Demonstration Research Corporation.

Wooldridge, J. (2002): “Econometric analysis of cross section and panel data”, Massachusetts Institute of Technology.

# Acronyms and Abbreviations

## Technical Terms

ATE	Average treatment effect
ATT	Average treatment effect on the treated
CIA	Conditional Independence Assumption
DD	Differences-in-differences estimator
IV	Instrumental variables estimator
OLS	Ordinary least squares estimator
OVB	Omitted variables bias
PSM	Propensity-score matching
SSR	Sum of squared residuals
SUTVA	Stable unit treatment value assumption

## Data sets and variable names

ABE	Adult Basic Education
AFDC	Aid to Families with Dependent Children
ESL	English as a Second Language
GED	General Educational Development certificate

## Study and program names

HCD	Human Capital Development approach
JOBS	Job Opportunities and Basic Skills Training

LFA	Labor Force Attachment approach
NEWWS	National Evaluation of Welfare-to-Work Strategies
TANF	Transitional Assistance for Needy Families

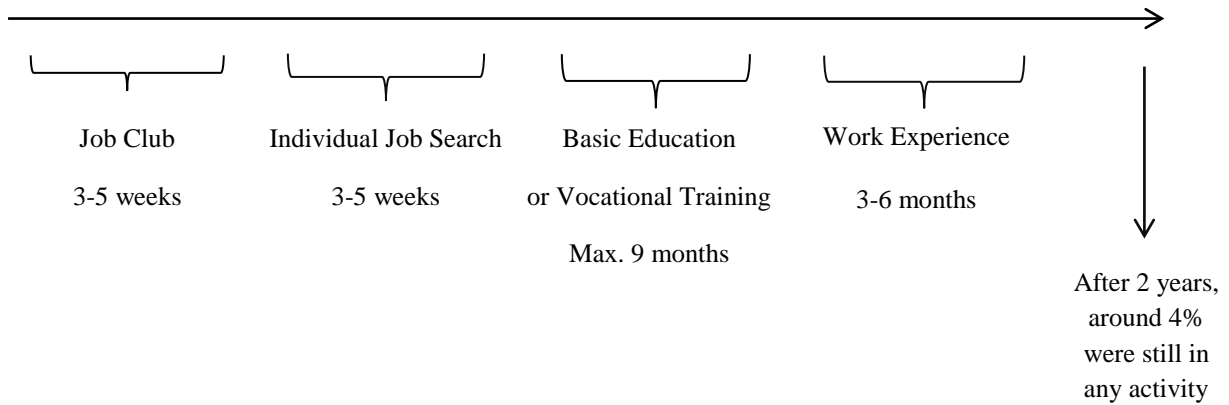
### **Others**

CASAS	Comprehensive Adult Student Assessment Systems
FSA	Family Support Act
TCC	Transitional Child Care
TMA	Transitional Medicaid Assistance

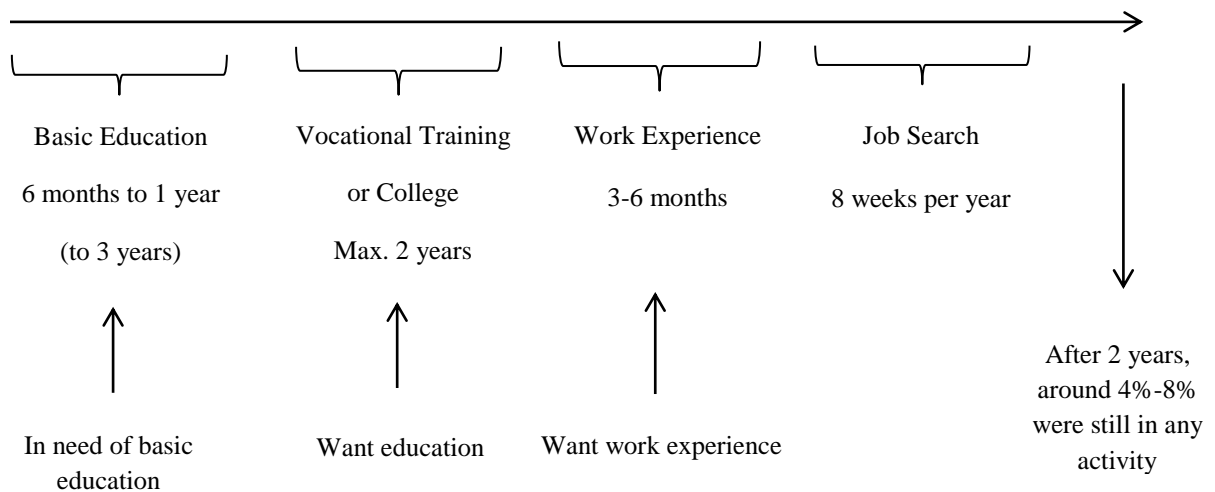
# **Appendix A**

## **Supplementary Tables to Section 2**

Appendix Figure A.1.- Labor Force Attachment Activities Sequence



Appendix Figure A.2.- Human Capital Development Activities Sequence





Appendix Table A.1.- Main characteristics of the training programs by Site

	Grand Rapids	Portland	Riverside
<b>1 Treatments</b>	HCD and LFA	Employment-focused approach	HCD and LFA
<b>2 Assignment</b>	<p>Three-way random assignment (2 program groups and 1 control group)</p> <p>AFDC applicants or recipients have a routine meeting with their income maintenance worker, where is determined if the individual is required to enroll in JOBS. Those participants determined to be JOBS-mandatory attend a JOBS orientation where the random assignment takes place.</p>	<p>Two-way random assignment (1 program group and 1 control group)</p> <p>Approved AFDC applicants and recipients were assigned to one of two groups. A program group (64%), eligible for JOBS services and subject to program participation requirements and a control group, not eligible for JOBS services and not subject to participation requirements.</p>	<p>Three-way random assignment</p> <p>AFDC applicants or recipients have a routine meeting with their income maintenance worker, where is determined if the individual is required to enroll in JOBS. Those participants determined to be JOBS-mandatory attend a JOBS orientation where they are evaluated to determine whether they require basic education. Those who had a high school diploma or GED, or scored 215 or above on both the math and literacy sections of the GAIN Appraisal test, and were proficient in English were randomly assigned to the LFA or control group. Those without a high school diploma or GED, who scored below 215 on either section of the GAIN test, or who required English remediation, were randomly assigned to the HCD, LFA and control group.</p>
<b>3 Sample characteristics</b>	<p>Individuals aged 18 or older</p> <p>Single-parent heads of AFDC cases</p> <p>96% were female and average age was 31</p> <p>50% were white and 40% were African-American</p> <p>Program coverage to parents with children as young as age 1</p>	<p>Individuals aged 21 or older</p> <p>Single-parent heads of AFDC cases</p> <p>93% were female and average age was 30</p> <p>Majority of the sample was white</p> <p>Program coverage to parents with children as young as age 1</p>	<p>Individuals aged 18 or older</p> <p>Single-parent heads of AFDC cases</p> <p>90% were female and average age was 31</p> <p>50% were white, 29% were Hispanic and 17% were African-American</p> <p>Parents exempted whose youngest child was under age 3</p>
<b>4 Enforcement and Sanctioning</b>			
a) Enforcement	High enforcement	High enforcement	High enforcement
b) Sanctioning	High sanctioning	Moderate sanctioning	Moderate sanctioning
Sanctioning characteristics	<p>Federal JOBS regulations governed the rule enforcement and sanctioning process in JOBS programs nationwide. The penalty for noncompliance was removal of the JOBS-mandatory client from the AFDC grant. For example, if an AFDC case consisted of a JOBS-mandatory parent with two children, and the parent failed to participate, the AFDC grant was reduced so that only the two children were covered. Sanctions were to continue until the sanctioned individual complied with the JOBS participation mandate, with a minimum sanction length of one month for the first "offense", three months for the second, and six months for the third and subsequent offenses.</p> <p>For a three-person family in 1993, a sanction would have resulted in a \$88 decrease in a monthly grant of \$474.</p>	<p>For a three-person family in 1993, a sanction would have resulted in a \$142 decrease in a monthly grant of \$460.</p>	<p>For a three-person family in 1993, a sanction would have resulted in a \$120 decrease in a monthly grant of \$624.</p>

Appendix Table A.1 (continued)

Grand Rapids			Portland		Riverside	
	HCD	LFA		HCD	LFA	
Referred for sanction <sup>b</sup> (%)	40,6	50,2	31,9	27,7	27,0	
Sanction imposed <sup>b</sup> (%)	37,6	41,5	20,5	14,9	8,7	
Average number of months in which sanction was in effect <sup>b</sup>	9,8	11,6	5,4	8,3	4,9	
Participants who say they were informed about penalties for noncompliance <sup>a</sup> (%)	82,4	80,9	67,6	71,9	69,5	
<b>5 Support services</b>						
Child care	JOBS legislation required states to guarantee child care to each participant with dependent children if child care was necessary for the client to attend a program activity or accept employment.					
Types of child care	The JOBS program paid for four major types of child care: care provided by nonimmediate relatives aged 16 or over; family day care, provided in a private residence for not more than 6 children; group home care, provided in a private residence for 7 to 12 children; or center-based care, provided in a nonresidential facility, typically for 13 children or more. Center-based care and family or group home care may be licensed by state social services agencies.					
	Child care assistance was first handled by a local child care coordinating council (prior to June 1993) and later by a special child care unit in the welfare office.	Child care needs had to be authorized by the client's integrated case manager at the welfare office. Clients were free to choose the type of provider and the specific provider.	Child care and support services payments were authorized by JOBS case managers.			
Reimbursement	Provided payments to licensed and unlicensed providers. Providers were reimbursed on an hourly basis. The Grand Rapids scale offered a maximum of \$1.50 for child care provided in the client's home; \$2.00 for child care in the provider's home; \$2.10 for care of children aged 2 and over in a licensed center; and \$2.65 for care of children under age 2 in a licensed center.	The maximum allowable payment was based on the age of the child, the kind of care given, and the type of providers. In 1992, the maximum rate paid for care by a relative or in family day care or group home care was \$2.00 an hour or \$371.00 a month per child. The maximum rate paid for a care in a child care center was \$3.50 an hour or \$450.00 a month per child.	Provided payments to licensed and unlicensed providers. Riverside's reimbursement rates were calculated by the hour, varied by children's ages, region of the county, and full-time or part-time care. The full-time rates for children between ages 2 and 5 were \$2.15 for unlicensed in-home care, \$2.23 for family day care homes, and \$2.93 for child care centers.			

## NOTES:

<sup>a</sup> MDRC calculations from the Two-Year Client Survey<sup>b</sup> MDRC calculations from MDRC-collected JOBS case file data

Appendix Table A.2.- Main characteristics of the training programs

	Education-focused approach <b>HCD</b>		Employment-focused approach <b>LFA</b> <b>Varied first activity</b>		
	Grand Rapids	Riverside	Grand Rapids	Riverside	Portland
<b>1 Goal</b>	Prepare people for jobs that offer sufficient wages and benefits to get them off and keep them off welfare.		Employment was the primary goal		
<b>2 Focus</b>	Welfare recipients should upgrade their skills before seeking work through basic education or vocational training. By investing more program resources upfront recipients will experience a bigger payoff in job quality and stability in the future.		Emphasized quick exposure to and entry into the labor market as the best route to earnings increases, job advancement and self-sufficiency. The LFA theory is that welfare recipients can best build their work habits and skills in the workplace and move up to better positions, even if their initial jobs are not high-paying or particularly desirable.		Emphasized full-time jobs that paid more than minimum wage, included benefits and offered room for advancement. Heavy focus on job development and placement activities.
<b>3 Strategy</b>	The HCD program begins with longer-term education and training, generally lasting up to two years. Job search and vocational training activities may be assigned if clients do not find employment through their education and training program or on their own initiative.		The LFA program begins with job search activities, followed by short-term education and training only for those unable to find employment during job search. If necessary, use the first job as a steppingstone to a better work opportunity.		Mixed services strategy: job search, education and training, and work experience activities. The program was a blend of strong LFA elements and moderate HCD elements.
	36% of the clients in Grand Rapids were assigned to basic education first.	Only those without a high school diploma or GED were included in the HCD group, 57% were assigned to basic education.	86% of the clients in Grand Rapids were assigned to job search first.	68% of the clients in Riverside were assigned to job search first.	Majority of clients were assigned to job search (fast track).
<b>4 Program activity components, sequence and emphasis</b>					
a) Activity sequence and emphasis	The starting point was an orientation that occurred immediately following random assignment. Orientation was where clients were informed about their program group, and it was where clients began to receive guidance from staff about what they would do next in JOBS.		Program group members attended a group JOBS orientation immediately following skills testing. Clients were selected by managers to attend 2 service tracks (selection was based on a variety of factors: employment history, educational status and personal goals). Fast track: clients ready to look for work (job club + job search)		
	Clients were encouraged to invest time in education or vocational training in order to prepare themselves for good jobs.		The first activity was usually job club, and the instruction and resources clients found there were uniformly designed to help them obtain rapid employment. Clients who did not obtain work after job club were usually assigned to short-term education, training, or unpaid work activities so that they could boost their skills somewhat and resume their job search as soon as possible.		

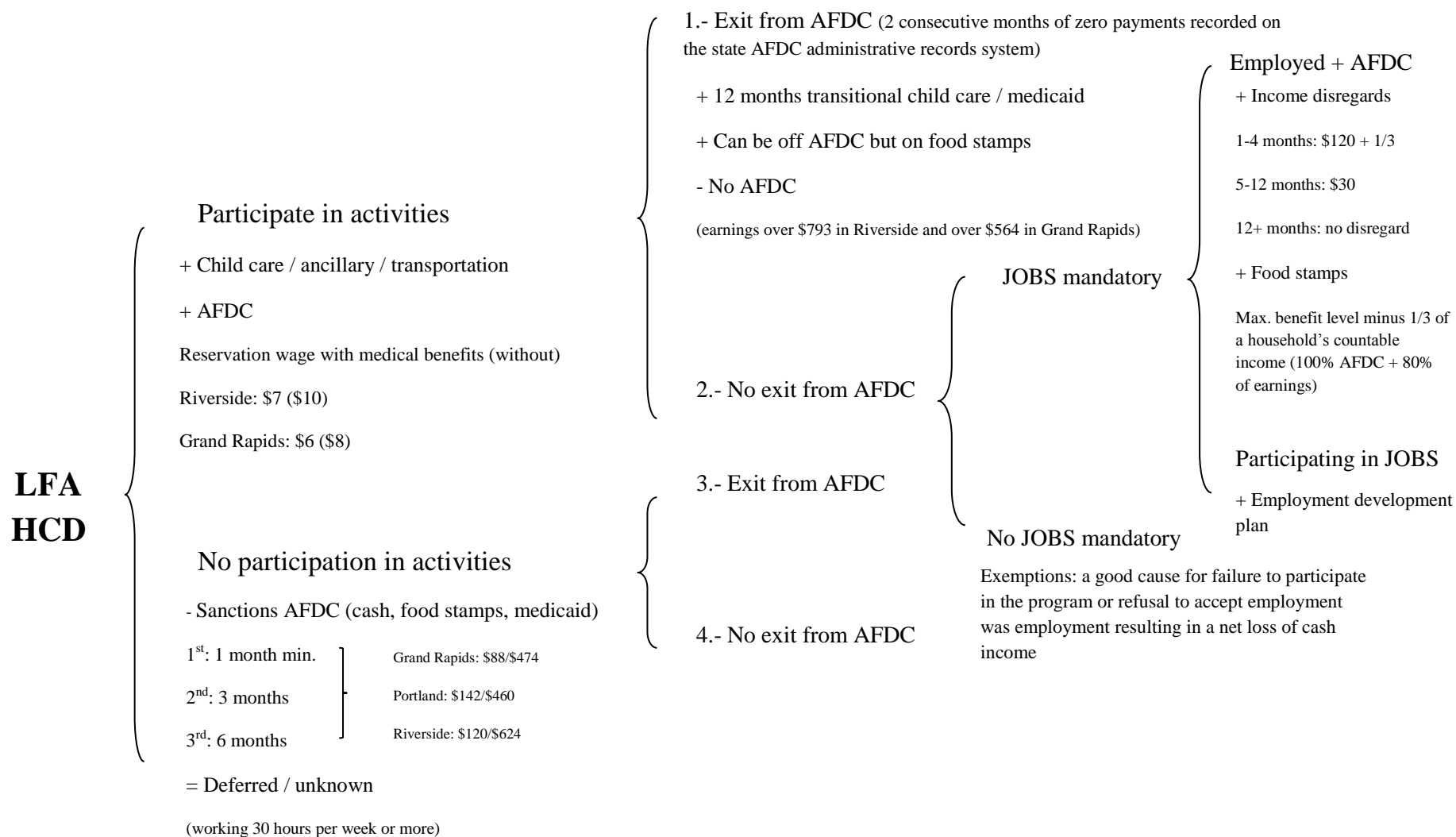
Appendix Table A.2 (continued)

	HCD		LFA		Varied first activity
	Grand Rapids	Riverside	Grand Rapids	Riverside	Portland
	In Grand Rapids, clients were encouraged to build up their reading, math, and vocational skills through basic education, vocational training or college. Basic education often meant high school completion classes rather than GED classes. Problem-solving skills applicable to the workplace were incorporated into the basic education curriculum. Small classes with whole-class instruction and individualized methods. Greater use of vocational training.	Riverside's program was restricted to clients who lacked a high school diploma or GED certificate. Thus, basic education was almost exclusively the first activity. Classes were larger and more likely to use computer-assisted instruction. Organizational context was extremely employment-focused.		Riverside was the only program that actively developed jobs and referred clients to employers. Staff in Riverside at every level believed strongly in the importance of getting clients to work.	Enhanced track: clients not ready to enter the labor market (life skills training class + basic education classes)
b) Content of program activities	Job club lasted between 1 week (Riverside, including an in-depth comparison of welfare and earned income) and 2 weeks (Grand Rapids, including career exploration). Clients attended these classes from 15 to 30 hours per week. Instructors tried to use classrooms and clients were told to come dressed as they would for a job and to show up on time. Clients were taught how to find job leads and complete job applications, how to conduct a successful interview, how to prepare a resume and cover letter, and how to identify and value their strengths and talents. The phone room segment immediately followed the classroom portion and the goal was to have clients apply their job-seeking skills by calling employers, arranging interviews and submitting job applications.	Individual job search: required clients to look for employment on their own, document the names of the employers they contacted, and report to a JOBS staff member each week on their progress. Number of employer contacts required of Grand Rapids and Riverside clients was determined on an individual basis by program staff. It was used for clients who had completed job club without finding work. Length of job club and individual job search was eight weeks maximum per year.			Job Club / Job Search: Classes 30 hours per week for two weeks. Discussion of career goals, resume preparation, videotaped practice interviews.
	Basic education was provided to clients who lacked a high school diploma or GED certificate (or, in Riverside, clients who possessed these credentials but with low scores). There were 4 major types of classes: high school completion, General Educational Development (GED), Adult Basic Education (ABE) and English as a Second Language (ESL).				Life skills training: Four- to five-week class involved examination of work history and vocational interest. Goal was to prepare people for work and eventual self-sufficiency. Basic education: six-week General Educational Development (GED) class for people who do not have a high school diploma. Adult Basic Education (ABE) for individuals whose achievements are lower than required.

Appendix Table A.2 (continued)

	HCD		LFA		Varied first activity
	Grand Rapids	Riverside	Grand Rapids	Riverside	Portland
	<p>Vocational training: the most common training programs included automotive maintenance and repair, business and clerical occupations, cabinet and furniture making, computer programming, cosmetology, electronics, nursing, refrigerator repair, and truck driving. Two-year programs led to an associate's degree.</p> <p>College: limited to clients who could complete an associate's or bachelor's degree within two years.</p> <p>Work experience: included 3 types of positions: unpaid work in the public or private nonprofit sectors; on-the-job training in the private sector, usually offering a wage subsidized by the client's welfare grant; and paid work, usually in the form of college work-study positions. Unpaid work experience was more common than on-the-job training or paid work. Clients' assignments usually lasted either three or six months.</p>		<p>Vocational training: short programs generally led to a certificate of credit.</p>		<p>Vocational training</p> <p>College</p> <p>Work experience: unpaid work in the non-profit and private sector, on-the-job training in the private sector and paid work. Participation was voluntary in unpaid positions, and positions lasted a maximum of 3 months.</p>
<b>5 Impacts</b>	<p>In total, over two years of follow-up, earnings gains matched or exceeded AFDC reductions for HCD high school graduates but were smaller than AFDC reductions for HCDs without a diploma or GED.</p>		<p>Two-year reductions in AFDC payments exceeded earnings gains in both Grand Rapids and Riverside.</p>		<p>The net increase in combined income of the treatment group relative to control members totaled just \$191 above the control group mean. Quarterly trends, however, show that program group members may improve their financial situation in the future.</p>

Appendix Figure A.3.- Labor Force Attachment and Human Capital Development Participation Patterns



Appendix Table A.3.- Main characteristics by program, treatment group versus control group

	Education-focused approach		Employment-focused approach			Control Group			
	HDC		LFA		Mixed-strategy				
	Grand Rapids	Riverside	Grand Rapids	Riverside	Portland	Grand Rapids	Riverside LFA	Riverside HDC	Portland
<b>1 Participation and Sanctioning <sup>a</sup></b>									
a) Participated in (%):									
Any activity	68,7	67,8	55,8	54,9	63,9	41,1	30,1	27,3	37,1
Job search	20,2	28,2	30,1	38,1	40,4	6,1	6,2	6,4	8,2
Education or training activity	59,2	58,6	33,7	23,6	39,1	37,3	25,1	23,3	29,2
Work experience or on-the-job training	3,8	2,1	4,2	2,7	9,4	1,6	2,0	1,2	2,3
b) Sanctioning (%)									
Sanctioned	32,3	22,1	35,1	15,2	18,4	6,7	3,9	4,2	4,4
<b>2 Estimated gross cost per group member within 2 years after orientation (in 1993 US\$) <sup>c</sup></b>									
Orientation and assesment	297	107	24	106	134	0	0	0	0
Job search	260	655	740	788	421	70	40	39	42
Basic education	1855	2103	736	155	488	713	95	163	303
Vocational training and college	2991	379	2403	848	1221	2297	596	367	1042
Work experience	191	57	110	50	222	11	58	27	33
Child care	542	164	366	88	1422	207	29	15	565
Total	6170	3540	4406	2082	4027	3298	819	609	2010
<b>3 Impacts of the program <sup>d</sup></b>									
a) Impacts on welfare status (\$) average total payments received years 1-2									
AFDC payments	6813	9235	6301	8385	5818	7639	9652	10369	7014
Food Stamp payments					3935				4391
b) Impacts on average total measured income (\$) Years 1-2									
Total earnings	4502	3278	4935	5386	7133	3916	4174	3090	5291
Combined income <sup>e</sup>					16886				16696

NOTES:

<sup>a</sup> MDRC calculations from the Two-Year Client Survey

<sup>b</sup> MDRC calculations from MDRC-collected JOBS case file data

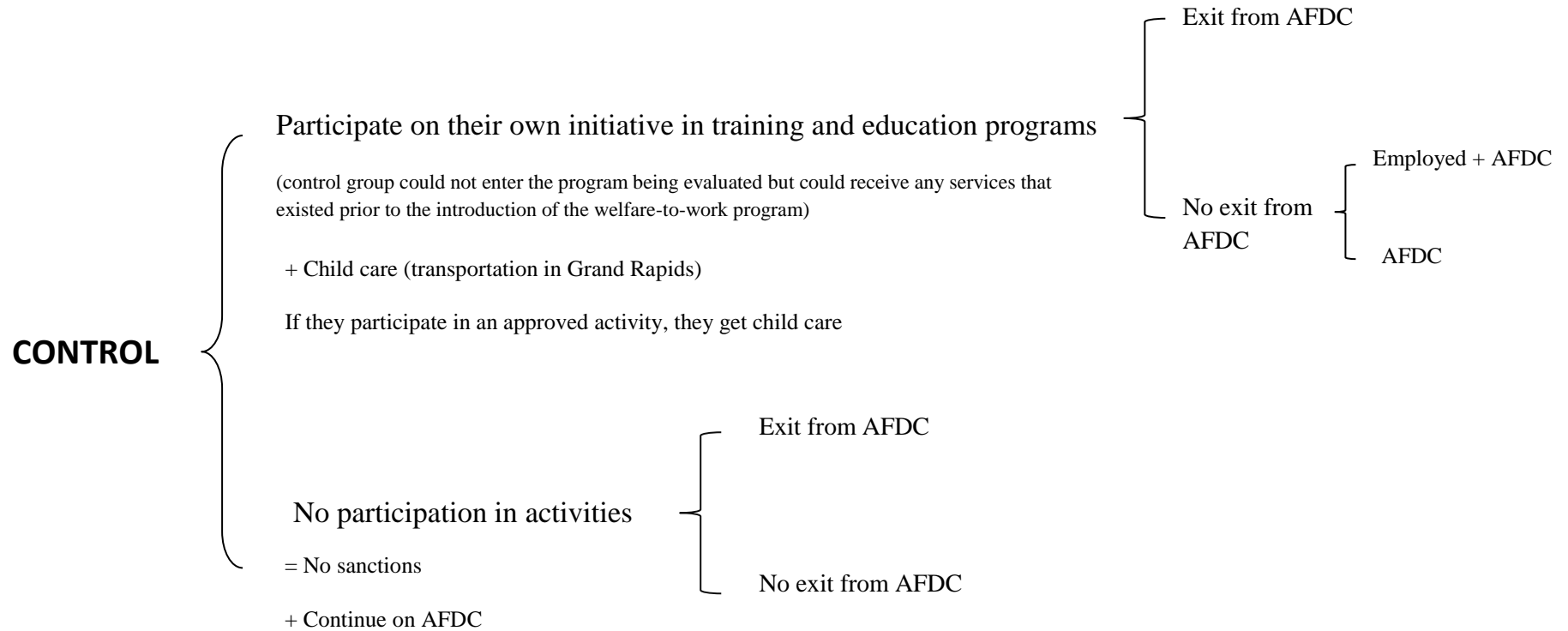
<sup>c</sup> MDRC calculations based on fiscal and participation data from Grand Rapids, Riverside and Oregon; in all three sites collected JOBS case file data and the MDRC Two-Year Client Survey.

MDRC child care calculations from Fulton County, Michigan, and Riverside County payment data. Other support service data from county records.

<sup>d</sup> MDRC calculations from Michigan, California and Oregon unemployment insurance (UI) earnings records and AFDC records.

<sup>e</sup> "Combined income" is income from earnings, AFDC, and Food Stamps.

Appendix Figure A.4.- Control Group Participation Patterns





# **Appendix B**

**Supplementary tables to Section 4**

Appendix Table B.1.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Grand Rapids.

Characteristic	Grand Rapids				
	Control	LFA	HCD	LFA-Control	HCD-Control
<b>Earnings, welfare payments and employment</b>					
Employed in year prior randomization	0,4914 (0,5001)	0,4846 (0,4999)	0,5054 (0,5001)	-0,0068 (0,0186)	0,0141 (0,0187)
Earnings in year prior randomization	1610,2160 (3531,5300)	1434,3620 (2901,9900)	1714,0240 (3901,5420)	-175,8534 (120,1202)	103,8086 (139,2868)
Received AFDC during year prior randomization	0,8403 (0,3665)	0,8215 (0,3831)	0,8130 (0,3900)	-0,0188 (0,0140)	-0,0273* (0,0142)
Monthly AFDC received in year prior to RA	341,1914 (184,5519)	326,9819 (186,5535)	322,1477 (186,7819)	-14,2095** (6,9207)	-19,0437*** (6,9408)
Number of months received AFDC year prior RA	7,8640 (4,8720)	7,7973 (4,9538)	7,6816 (5,0047)	-0,0667 (0,1833)	-0,1825 (0,1847)
Received Food Stamps in year prior RA	0,8892 (0,3140)	0,8718 (0,3344)	0,8611 (0,3459)	-0,0174 (0,0121)	-0,0281** (0,0124)
Monthly Food Stamps received	188,8885 (90,6388)	183,4262 (94,1922)	181,8963 (96,1581)	-5,4623 (3,4491)	-6,9921** (3,4954)
Number of months receiving Food Stamps prior random assignment	8,3791 (4,5547)	8,3175 (4,7257)	8,1782 (4,7640)	-0,0617 (0,1732)	-0,2010 (0,1743)
<b>Demographic Characteristics</b>					
Single parent, ever married	0,4176 (0,4933)	0,4013 (0,4903)	0,4027 (0,4906)	-0,0162 (0,0183)	-0,0149 (0,0184)
One child	0,4561 (0,4982)	0,4741 (0,4995)	0,4586 (0,4985)	0,0180 (0,0186)	0,0025 (0,0186)
Two children	0,3583 (0,4797)	0,3465 (0,4760)	0,3602 (0,4802)	-0,0117 (0,0178)	0,0020 (0,0179)
Three or more children	0,1856 (0,3889)	0,1793 (0,3837)	0,1811 (0,3853)	-0,0063 (0,0144)	-0,0045 (0,0145)
Any child 0-5 years old	0,6921 (0,4618)	0,6971 (0,4597)	0,6832 (0,4654)	0,0050 (0,0172)	-0,0089 (0,0173)
Black	0,4086 (0,4918)	0,4047 (0,4910)	0,3726 (0,4837)	-0,0039 (0,0183)	-0,0360** (0,0182)
Not black or white	0,1101 (0,3131)	0,1074 (0,3097)	0,1064 (0,3084)	-0,0027 (0,0116)	-0,0037 (0,0116)
Age at random assignment	29,6547 (6,4833)	29,5772 (6,4624)	29,8171 (6,5897)	-0,0775 (0,2414)	0,1624 (0,2444)
High school diploma or GED	0,5928 (0,4915)	0,5779 (0,4941)	0,6003 (0,4900)	-0,0150 (0,0184)	0,0075 (0,0183)
<b>Sample size</b>	1390	1490	1476		

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

Appendix Table B.2.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Riverside in need of education.

Characteristic	Riverside in need of education				
	Control	LFA	HCD	LFA-Control	HCD-Control
<b>Earnings, welfare payments and employment</b>					
Employed in year prior randomization	0,3186 (0,4661)	0,3300 (0,4704)	0,3406 (0,4741)	0,0114 (0,0177)	0,0219 (0,0177)
Earnings in year prior randomization	1293,9310 (3173,1940)	1434,4850 (3521,4580)	1477,9020 (3479,3890)	140,5546 (126,7320)	183,9714 (125,6527)
Received AFDC during year prior randomization	0,7731 (0,4190)	0,7680 (0,4223)	0,7874 (0,4093)	-0,0051 (0,0159)	0,0143 (0,0156)
Monthly AFDC received in year prior to RA	440,8165 (306,1994)	439,0804 (306,5155)	444,4448 (296,7973)	-1,7361 (11,5760)	3,6283 (11,3672)
Number of months received AFDC year prior RA	5,9588 (5,1045)	5,9908 (5,1649)	6,1573 (5,1302)	0,0320 (0,1940)	0,1985 (0,1930)
Received Food Stamps in year prior RA	0,6893 (0,4629)	0,6742 (0,4688)	0,6762 (0,4681)	-0,0151 (0,0176)	-0,0131 (0,0176)
Monthly Food Stamps received	120,9689 (107,2307)	120,1403 (112,2407)	118,6112 (106,5796)	-0,8286 (4,1487)	-2,3577 (4,0309)
Number of months receiving Food Stamps prior random assignment	4,7868 (4,8399)	4,8759 (4,9014)	5,0000 (4,9504)	0,0890 (0,1841)	0,2132 (0,1846)
<b>Demographic Characteristics</b>					
Single parent, ever married	0,6344 (0,4818)	0,6438 (0,4790)	0,6478 (0,4778)	0,0094 (0,0182)	0,0134 (0,0181)
One child	0,3543 (0,4785)	0,3495 (0,4770)	0,3574 (0,4794)	-0,0048 (0,0181)	0,0031 (0,0182)
Two children	0,3090 (0,4622)	0,3004 (0,4586)	0,3213 (0,4671)	-0,0086 (0,0175)	0,0123 (0,0176)
Three or more children	0,3367 (0,4728)	0,3502 (0,4772)	0,3213 (0,4671)	0,0134 (0,0180)	-0,0154 (0,0178)
Any child 0-5 years old	0,5917 (0,4917)	0,5794 (0,4938)	0,5768 (0,4942)	-0,0123 (0,0187)	-0,0149 (0,0187)
Black	0,1561 (0,3631)	0,1650 (0,3713)	0,1601 (0,3669)	0,0090 (0,0139)	0,0041 (0,0138)
Not black or white	0,4415 (0,4967)	0,4394 (0,4965)	0,4601 (0,4986)	-0,0021 (0,0188)	0,0187 (0,0188)
Age at random assignment	32,1676 (6,9445)	32,4260 (6,9123)	32,6014 (7,2866)	0,2583 (0,2618)	0,4338 (0,2685)
High school diploma or GED	0,2262 (0,4185)	0,2348 (0,4240)	0,2161 (0,4117)	0,0087 (0,0159)	-0,0101 (0,0157)
<b>Sample size</b>	1384	1418	1430		

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

Appendix Table B.3.- Selected characteristics of female sample members with two years of earnings data prior to random assignment, Riverside not in need of education.

Characteristic	Riverside not in need of education		
	Control	LFA	LFA-Control
<b>Earnings, welfare payments and employment</b>			
Employed in year prior randomization	0,4581 (0,4984)	0,4542 (0,4981)	-0,0039 (0,0177)
Earnings in year prior randomization	2645,8760 (5016,3230)	2245,3570 (4169,1280)	-400,5188** (164,0885)
Received AFDC during year prior randomization	0,7481 (0,4342)	0,7745 (0,4181)	0,0264* (0,0152)
Monthly AFDC received in year prior to RA	402,2361 (290,1048)	403,6930 (281,7797)	1,4569 (10,1758)
Number of months received AFDC year prior RA	5,4308 (5,0441)	5,4864 (5,0389)	0,0556 (0,1794)
Received Food Stamps in year prior RA	0,6516 (0,4766)	0,6557 (0,4753)	0,0041 (0,0169)
Monthly Food Stamps received	110,4156 (103,6211)	111,5666 (104,0292)	1,1510 (3,6946)
Number of months receiving Food Stamps prior random assignment	4,4099 (4,7590)	4,4542 (4,8514)	0,0443 (0,1710)
<b>Demographic Characteristics</b>			
Single parent, ever married	0,6815 (0,4660)	0,6865 (0,4640)	0,0050 (0,0166)
One child	0,4180 (0,4934)	0,4310 (0,4954)	0,0130 (0,0177)
Two children	0,3402 (0,4739)	0,3416 (0,4744)	0,0014 (0,0170)
Three or more children	0,2418 (0,4283)	0,2273 (0,4192)	-0,0145 (0,0152)
Any child 0-5 years old	0,5711 (0,4951)	0,5619 (0,4963)	-0,0091 (0,0177)
Black	0,1802 (0,3845)	0,1718 (0,3773)	-0,0084 (0,0136)
Not black or white	0,2126 (0,4093)	0,2135 (0,4099)	0,0010 (0,0146)
Age at random assignment	32,0558 (6,7865)	32,1636 (6,7352)	0,1078 (0,2406)
High school diploma or GED	0,9987 (0,0356)	0,9987 (0,0355)	0,0000 (0,0013)
<b>Sample size</b>	1576	1583	

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

# **Appendix C**

**Supplementary tables to Section 6**

Appendix Table C.1.- Comparison between short-run and medium-run impacts measured using experimental and non-experimental data: Grand Rapids

Treatment and control group		Selection on observables						Selection on unobservables	
		Difference of Means		OLS Regression		Propensity-score subclassification		Differences-in-differences	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Short Run	Medium Run	Short Run	Medium Run	Short Run	Medium Run	Short Run	Medium Run
Grand Rapids (control)	LFA	466,63***	80,65						
	HCD	431,41**	228,10						
Grand Rapids (Riverside)	LFA	610,07	1707,17	939,82	2086,30	747,70	1859,83	1152,40	2246,93
	HCD	574,85	1854,62	904,60	2233,75	712,48	2007,28	1117,18	2394,38
Grand Rapids (Portland)	LFA	309,12	321,19	233,63	364,10	237,99	327,63	15,97	24,40
	HCD	273,90	468,64	198,41	511,55	202,77	475,08	-19,25	171,85

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Note: Short run is defined as the two years following random assignment and medium run is defined as the third through fifth years following random assignment

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.

Appendix Table C.2.- Comparison between short-run and medium-run impacts measured using experimental and non-experimental data: Riverside

Treatment and control group		Selection on observables						Selection on unobservables	
		Difference of Means		OLS Regression		Propensity-score subclassification		Differences-in-differences	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Short Run	Medium Run	Short Run	Medium Run	Short Run	Medium Run	Short Run	Medium Run
Riverside (control)									
In need of education	LFA	758,37***	453,85**						
	HCD	323,92**	317,03						
Not in need of education	LFA	863,51***	230,82						
Riverside (Portland)									
In need of education	LFA	494,09	-888,20	171,90	1209,03	263,03	-1078,67	-378,07	-1768,67
	HCD	59,64	-1025,02	-262,55	-1345,85	-171,42	-1215,46	-812,52	-1905,49
Not in need of education	LFA	599,23	-1111,23	277,04	-1432,06	368,17	-1301,67	-272,93	-1991,7
Riverside (Grand Rapids)									
In need of education	LFA	614,93	-1172,67	285,18	-1551,80	477,30	-1325,33	72,60	-1712,43
	HCD	180,48	-1309,49	-149,27	-1688,62	42,85	-1462,15	-361,85	-1849,25
Not in need of education	LFA	720,07	-1395,70	390,32	-1774,83	582,44	-1548,36	177,74	-1935,46

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Note: Short run is defined as the two years following random assignment and medium run is defined as the third through fifth years following random assignment

Standard errors are shown in parentheses.

All dollar amounts are in 1996 dollars.